Q. J. R. Meteorol. Soc. (2002), **128**, pp. 285–321

Influence of added observations on analysis and forecast errors: Results from idealized systems

By R. E. MORSS* and K. A. EMANUEL
Massachusetts Institute of Technology, USA

(Received 4 August 1999; revised 31 May 2001)

SUMMARY

Recent tests of adding observations to improve individual forecasts have shown a mix of positive, negative, and neutral results. In this study, the influence of added observations on analysis and forecast errors is explored in idealized systems. The primary system studied is a three-dimensional simulated observing and forecasting system that includes a quasi-geostrophic forecast model and a three-dimensional variational data-assimilation system. In this simulated system, adding observations generally improves analyses and often improves 12-hour forecasts. Even with perfect observational data and a perfect forecast model, however, there is a non-negligible risk that adding observations will degrade analyses, and a significant risk that assimilated added observations will degrade forecasts on a time-scale of one or several days. To illustrate several general principles that help interpret the results, several experiments are performed with low-dimensional data-assimilation systems; the results demonstrate that the risk of analysis degradation is inherent in statistical data assimilation. Experiments are also performed with a low-dimensional forecast model, demonstrating that the risk of forecast degradation is inherent in prediction systems with sensitivity to small errors in initial conditions. Although degradations cannot be avoided, several circumstances are identified in which adding observations is less likely to degrade analyses and forecasts and on average improves analyses and forecasts by a larger amount.

KEYWORDS: Analysis and forecast degradations Data assimilation Targeted observations

1. INTRODUCTION

One of the primary causes of errors in numerical weather forecasts is errors in the analyses used as the initial conditions in the forecast model. More observations can provide more information about the true atmospheric state, and thus can help reduce such errors. At first glance, it seems obvious that more data should improve atmospheric analyses, as long as the data is reasonably accurate. Given a reasonably good forecast model, these improved initial conditions should, in turn, produce improved forecasts.

Recently, there has been interest in adding observations in potentially important locations, called targeting or adapting observations, to improve specific significant forecasts. Targeting observations has been tested for midlatitude-winter forecasts in the Fronts and Atlantic Storm-Track EXperiment, the North Pacific Experiment, and the Winter Storm Reconnaissance Program, and for tropical-cyclone forecasts. For sample results, see Emanuel and Langland (1998), Langland (1999), Bergot (1999), Gelaro *et al.* (1999), Langland *et al.* (1999a), Szunyogh *et al.* (1999), Langland *et al.* (1999b), and Szunyogh *et al.* (2000). Although the results include a limited number of cases and are preliminary, so far it is apparent that while adding observations in targeted regions has improved some of the forecasts, in other cases it has had little effect or has even degraded forecasts.

Adding observations could degrade deterministic forecasts in the real world for a number of reasons. The degradations could be caused by limitations in the observing system; for example, the observing platforms may have been unable to collect data in some of the desired locations, the measurements could be unrepresentative in space or in time, or the data quality could be unexpectedly poor. The degradations could be due to errors in the forecast model, or they could be an artefact of insufficient data for forecast verification. Adding observations could also degrade some forecasts because of how

* Corresponding author, present affiliation: National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307, USA. e-mail: morss@ucar.edu

the data-assimilation procedure incorporates the observations into the model state, and how the forecast model (even without model errors) subsequently uses the assimilated information.

In this study, we begin to investigate degradations caused by the last possibility, in the absence of observing-platform constraints and forecast-model errors. To do so, we explore the effects of adding observations in an idealized, but three-dimensional, simulated observing and forecasting system, using perfect observing platforms, a perfect forecast model, and perfect information for forecast verification. We also explore how data-assimilation systems and nonlinear forecast models use added information in a few simple, low-dimensional experiments.

Despite their simplicity relative to the real atmosphere, the systems studied exhibit a wide variety of complex behaviour, more than can be understood in a single study. In addition, many details of the results are specific to the data-assimilation systems, forecast models, and idealized experimental set-ups tested. Thus, we concentrate on several general aspects of the results that are likely to be relevant in a variety of numerical weather-forecasting systems. Although the results and discussion raise questions to be addressed in future work, they are an important step towards understanding how observations, data-assimilation systems, and numerical forecasts interact.

The results show that adding observations improves many analyses and forecasts. Even with good observations and a good forecast model, however, added observations sometimes degrade analyses, and they frequently degrade longer-term forecasts. Analyses can be degraded for two reasons. First, as is well known, the statistics used in current data-assimilation systems are imperfect. Second, data assimilation is a statistical procedure, with an inherent risk of degradation. Forecasts can be degraded not only because the analyses used as the initial conditions can be degraded, but also because error growth limits predictability. The possibility of analysis and forecast degradations follows from data-assimilation theory and predictability theory. The implications of degradations for targeting observations, however, are not widely appreciated.

The results also identify circumstances in which adding observations is less likely to degrade analyses and forecasts and, on average, improves them more. This suggests that the risk of degradation can be decreased if, when selecting observation locations, one understands how additional information is likely to be used by the data-assimilation system and the forecast model in different situations. Several studies have speculated that poor data assimilation has limited the effectiveness of real-world targeted observations (see, for example, the references listed above for field-experiment results); thus, the circumstances discussed in this paper focus primarily on the role of data assimilation in degradations. The role of forecast-error growth in degradations is addressed only briefly and in general.

Section 2 describes the set-up of the simulated observing and forecasting system, including the quasi-geostrophic forecast model and the three-dimensional variational (3D-Var) data-assimilation system. Section 3 explains how the influence of adding observations is tested at each location in the simulated domain for a number of different forecast cases. Section 4 presents general results from these experiments for several sample cases and compiled over many cases. In section 5, we use idealized low-dimensional statistical data-assimilation systems to begin to investigate how and why incorporating added data with statistical data-assimilation systems can degrade analyses. In section 6, we use a three-variable Lorenz model to demonstrate that nonlinear error growth limits forecast improvement, even when the initial conditions are improved and the forecast model is perfect. Section 7 revisits the results from section 4, discussing several general circumstances in the simulated analysis and forecasting system in which

analysis and forecast improvements are more likely and the mean change is a larger improvement. Section 8 reviews the major results and conclusions.

2. Simulated observing and forecasting system

(*a*) *Quasi-geostrophic model*

The quasi-geostrophic (QG) model is a grid-point channel model on a beta plane; it is zonally periodic and has multiple vertical levels. The model was developed at the National Center for Atmospheric Research (NCAR) and is the same as that in Morss *et al.* (2001). The QG model has simplified geometry and time- and longitude-independent forcing, but otherwise exhibits baroclinic dynamics similar to those on synoptic scales in the real atmosphere. Its relative simplicity makes both performing many experiments feasible and the results easier to interpret. The model is described only briefly here; for further details, see Rotunno and Bao (1996) and Morss (1999).

The QG model is forced by relaxing the model variables (interior potential vorticity and upper- and lower-boundary potential temperature) to a specified zonal-mean state. The zonal-mean state is a Hoskins–West jet (Hoskins and West 1979) with a zonal-wind maximum of 60 m s^{-1} at the tropopause and a corresponding meridional temperature gradient; the relaxation time-scale is 20 days. The stratification is constant, and the tropopause height is fixed but with varying temperature. The model dissipation consists of fourth-order horizontal diffusion and, at the lower boundary, Ekman pumping.

The channel circumference is 16 000 km, the channel width is 8000 km, and the depth of the domain is 9 km. The Rossby radius of deformation is approximately 1000 km. The results shown here have 250 km horizontal grid spacing and five vertical levels. At this grid spacing, small errors double in magnitude in approximately 2–3.5 days. A sample state of the QG model is depicted in Fig. 1.

(*b*) *Data-assimilation system*

The data-assimilation system is a 3D-Var scheme based on the Spectral Statistical-Interpolation analysis system operational at the National Centers for Environmental Prediction in the USA (Parrish and Derber 1992). At each assimilation time, the 3D-Var scheme combines a weighted short-term model forecast (also called a background field) with weighted observations to create an analysis field, which is then used as the initial conditions for forecasts. The difference between the analysis and the background is called the analysis increment. The 3D-Var scheme is the same as that used in Morss *et al.* (2001) and is only summarized here; for further details, see also Morss (1999).

All observations in these experiments simulate rawinsondes that measure winds and potential temperature at model grid points and model levels. The weighting of the observations in the analysis is controlled by the observation-error statistics given to the 3D-Var scheme. The observation errors are assumed to be uncorrelated between different rawinsondes and between wind and temperature observations (Dey and Morone 1985), and the observation-error covariance matrix was developed from the values given in Parrish and Derber (1992) and from the simple function in Eq. (3.19) in Bergman (1979).

The background-error statistics given to the 3D-Var scheme help determine how the data assimilation weights different parts of the background field relative to the observations. They also determine how the data assimilation spreads information from observed to unobserved locations. In order for the inversion of the data-assimilation equation to be computationally feasible, the structure of the complete time- and space-varying background-error covariances must be simplified. Following Parrish and Derber (1992),

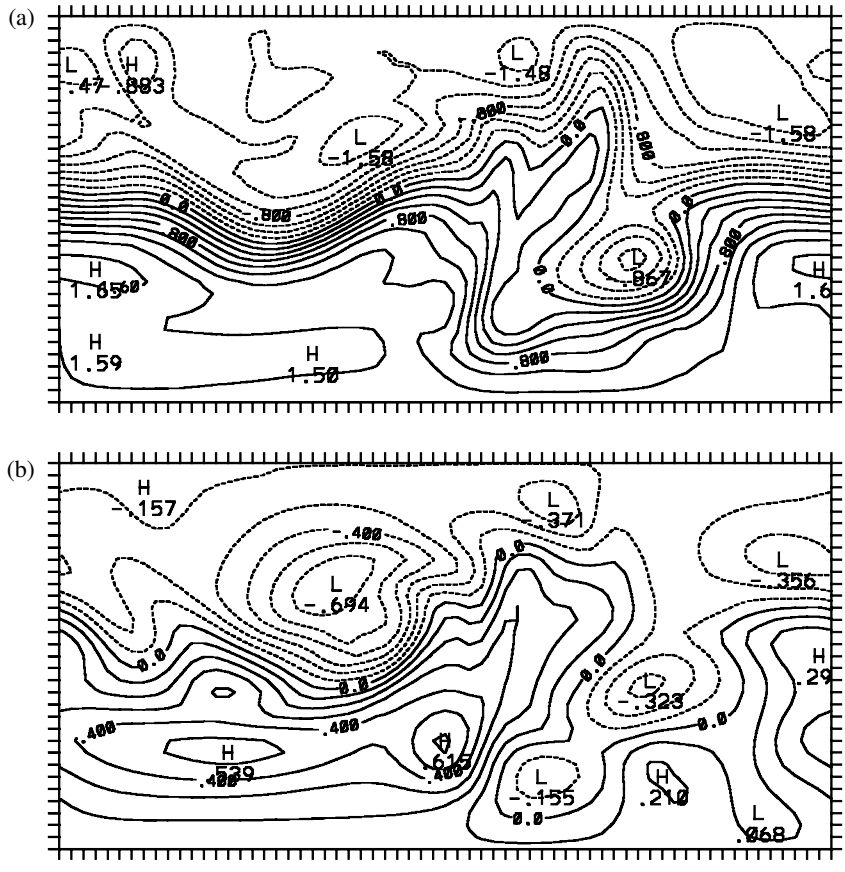

Figure 1. Synoptic situation (the true state of the quasi-geostrophic model) at day 0, the time at which added observations are tested in the examples shown in Figs. 3–7 and Tables 1–2. Stream function, in non-dimensional units, is plotted at the upper boundary (a) and at the lower boundary (b). The x-axis is longitude (periodic) and the y-axis is latitude. Each grid point is marked with a tick, spaced by 250 km.

we assume that the background-error covariance matrix is fixed in time, is diagonal in spectral space, and has separable vertical and horizontal structures. In addition, only simple vertical background-error covariances (in terms of interior potential vorticity and boundary potential temperature) are specified.

With these assumptions, the background-error statistics, and thus the analysis increments for individual observations, are nearly isotropic and have no knowledge of specific atmospheric situations. As described in Morss *et al.* (2001), the background-error covariance matrix was generated by accumulating 12-hour forecast-error statistics over a large number of runs with different distributions of 32 fixed simulated rawinsondes. Sample analysis increments for a single observation are depicted in Morss *et al.* (2001); the effective background-error correlation length for stream function is of the order of 1000 km.

Although more complex data-assimilation schemes are currently being developed and implemented, we chose a 3D-Var scheme for several reasons. First, given the technology and information available when the experiments were performed, assumptions similar to ours were necessary to develop a data assimilation that is reliable, yet computationally fast enough to allow many experiments. Second, the 3D-Var method used

in this study is similar to the schemes currently operational at most numerical weather-prediction centres, and is likely to remain so in the near future. In addition, the 3D-Var scheme demonstrates how adding observations can, in general, influence analyses and forecasts in a three-dimensional model with an imperfect statistical data-assimilation system.

(*c*) *Set-up of identical twin experiments*

The identical twin experiments (also called observing-system simulation experiments, or OSSEs, with a perfect forecast model) are set up as in Morss *et al.* (2001). First, an arbitrary state of the QG model is selected to be the initial 'true' state. To create an initial 'model' state, random uncorrelated noise is added to the true state, independently at each grid point. Then, we run an analysis and forecast cycle by integrating both the true and model states forward in time, interrupting the model run every 12 hours to assimilate simulated rawinsonde observations of the true state into the model state. Before the start of the experiments, the analysis and forecast cycle is run for a 60-day spin-up period, so that the model state is largely both independent of its initial perturbation and equilibrated to the observational network. All analyses and forecasts are evaluated with respect to the (known) true state.

The simulated rawinsonde observations are generated by sampling the winds and potential temperature of the true state. During the spin-up period, random errors consistent with the observation-error statistics in the 3D-Var scheme are added to the observations (as described in Houtekamer 1993). To prevent the results at any given location and time from being biased by an observation with an unusually large error, the Gaussian distribution used to generate the observation errors is truncated at one standard deviation.

Along with the control model state, a 96-member ensemble of equally likely initial conditions is generated. To produce each ensemble member, we begin with a different perturbation of the true state, then run the analysis and forecast cycle with the same observations as were used to produce the control model state, but resampling the distribution of random observation errors. This ensemble is similar to the perturbed observation (also called multiple replication) ensemble in Morss *et al.* (2001) (see also Houtekamer and Derome 1995; Lorenz and Emanuel 1998; Hamill *et al.* 2000), except we perturb the true (error-free) observations rather than the control observations (which already contain errors).

The same QG forecast equations, at the same resolution, are used to integrate both the true and the model states forward in time. This simulates perfect knowledge of the atmospheric dynamic equations. It also assumes that all processes in the true atmosphere are resolved in the model atmosphere. We assume a perfect forecast model so that forecast errors can result only from errors in the initial conditions; this both removes a possible source of forecast degradations and simplifies interpreting the results. In a more realistic system, with an imperfect forecast model, the results are almost certain to be less optimistic.

3. Set-up of added-observation experiments in the simulated system

(*a*) *Generation of cases*

The influence of added observations is tested in a number of different cases. A 'case' is defined by a unique set of a true state of the QG model (which we call the 'synoptic situation') and a model state (an estimate of the true state, also called the analysis or the initial conditions for a forecast). The difference between the model state and the true state is called the 'initial-condition errors'. Seventeen synoptic situations are tested,

each separated by 12 hours during an 8-day run. Because a synoptic situation can be paired with many different initial-condition errors, multiple cases can be generated for each synoptic situation.

The initial-condition errors for the cases are generated in two ways. First, we produce six sequences of model states by running the OSSE six times over the same 60-day spin-up period and 8-day sequence of synoptic situations, but using a different 'standard observation network' for each OSSE run. For the results shown, each standard observation network is a different set of 16 fixed locations with simulated rawinsondes; the locations are selected randomly as described in Morss *et al.* (2001). Second, for one of the standard observation networks, we randomly select nine sequences of equally likely model states from the 96 remaining members of the ensemble described in section 2(c). For one standard observation network, then, we include cases for ten realizations of the initial-condition errors; for the other five standard observation networks, we include cases for only one realization of the initial-condition errors. In all, adding observations is tested in 255 cases (15 initial-condition errors $\times$ 17 synoptic situations), each with the same density of standard observations.

We generated cases in this manner to evaluate how the influence of added observations depends on the synoptic situation, the past observation locations, and the specific realization of the initial-condition errors. A standard network of 16 fixed observations was chosen because, at this observation density, adding only a few observations at every data-assimilation interval can reduce the mean analysis and forecast error by a large amount (Morss *et al.* 2001). We also tested adding observations to fixed observation networks of other densities; the results are qualitatively similar, but are in general more pessimistic for higher-density networks and more optimistic for lower-density networks.

The 8-day sequence of synoptic situations, the six standard observation networks, and the standard observation network used to generate the equally likely initial-condition errors were chosen arbitrarily. Since the experiments performed (described in section 3(b)) are computationally very expensive, the primary factor limiting the number and type of cases included was computational resources.

(*b*) *Testing the influence of added observations*

In each case, we first calculate the global-average error in the standard analysis (with only observations from the standard network assimilated) and in the forecasts produced from the standard analysis. This is the control run. Then, we systematically test the effect of adding observations at each location in the domain. To do so, we assimilate each test set of added observations along with the standard observations, then run forecasts from each test analysis. Global-average errors are calculated for the resulting analyses and forecasts, evaluated with respect to the true state. For each test set of added observations, the change in error for the analysis and for each forecast is defined as:

$$\Delta \text{ error} = \frac{\text{error with added observations} - \text{error without added observations}}{\text{error without added observations}}. \quad (1)$$

All results in sections 4 and 7 use Eq. (1), the percent change in global-average error in a single realization, to measure the influence of each set of added observations*.

The results shown are for global-average energy errors, summed over all levels as well as over the horizontal domain. Qualitatively, results are similar for global

* This error norm, which compares deterministic forecasts made with added observations with those made without added observations, is patterned after other recent studies of the influence of added observations (see, for example, the references provided in the introduction).

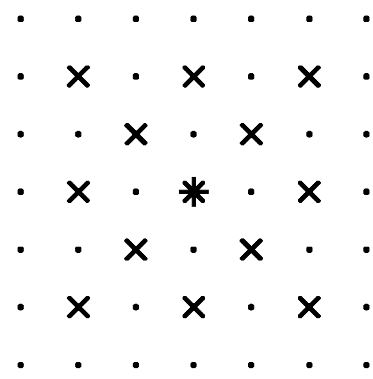

Figure 2. Observation pattern for cluster 13, the primary cluster used to test added observations. Dots represent grid points (spaced by 250 km), a + is plotted at the central location for the test observations, and ×'s are plotted at observed locations. The x-axis is longitude and the y-axis is latitude.

root-mean-square average potential-vorticity and stream-function errors, and for errors averaged over the horizontal domain but at individual levels. A domain-average error norm is used instead of a local error norm because it would be difficult to identify an appropriate local verification region for each of the analyses and forecasts tested. As discussed in Morss *et al.* (2001) and Morss (1999), in this simulated system, global-average analysis and forecast errors tend to be dominated by large errors in a few regions. Thus, a global-average error norm measures primarily the influence of added observations in the few regions with large errors.

With the definition in Eq. (1), negative values of error change indicate that the added test observations decrease the error, that is, improve the analysis or forecast. Positive values indicate that the added observations increase the error, that is, degrade the analysis or forecast. The effect of adding observations is tested at each location in the domain, except at the channel walls and at the locations where observations are already taken as part of the standard network; at these locations, the analysis- and forecast-error changes are set to zero.

In most of the results shown, the test observations are taken with a cluster of several simulated rawinsondes centred around each test location. We tested observation clusters to provide the 3D-Var data assimilation with more information about the structure of the initial-condition errors. The primary cluster used is 13 observations at alternating grid points in a 1000 km × 1000 km area, depicted in Fig. 2. Observations that extend outside the domain are removed; observations that overlap with standard observation locations are not removed. This cluster was selected because it was, on average, one of the more beneficial clusters in previous tests of adaptive observations with this simulated system (Morss 1999). Using it thus gives the data-assimilation system an advantage in improving analyses and forecasts. Single simulated rawinsondes and different-sized observation clusters were also tested for a subset of the cases; results are discussed briefly in section 4(a) and are presented in more detail in Morss (1999).

The results shown are for perfect (error free) test observations. We chose to present results for perfect observations so that degradations could not be attributed to errors in the observations. Note, however, that even with perfect observations, the data-assimilation system still assumes some observational error. Therefore, the analysis does not match the observations exactly. For a subset of the cases, the experiments were also

performed with test observations with errors; the results are similar to those shown and are discussed briefly in section 4(b).

Because it is computationally very expensive to assimilate test observations for each location in the domain and to generate forecasts for each test analysis, only a limited number of cases were tested. The cases are by no means a random sample; they include only 17 synoptic situations separated by only 12 hours, and only a few sets of initial-condition errors. Therefore, the results are not strictly representative of all possible cases. Nevertheless, given the large number of added sets of observations tested, the results suggest some ways that added observations can influence analysis and forecast errors.

4. General influence of added observations in the simulated system

In this section, we show results from the added-observation experiments in several sample cases, discussing the general similarities and differences. Then, we compile the results over all cases to show the overall likelihood of different-sized improvements and degradations in analyses and forecasts. Finally, we briefly discuss the change in error that is expected, averaged over the results for several equally likely realizations of initial-condition errors. Only general features of the results are discussed; several aspects are explored in further detail in section 7.

(*a*) *Sample results in individual cases*

Results are shown for three sample cases, with the same synoptic situation but different initial-condition errors. Figure 1 depicts the true synoptic situation at the observation time. For the first sample case, Fig. 3 depicts the spread in the 97-member ensemble of 12-hour forecasts valid at the observation time (an estimate of the initial-condition errors), and Fig. 4 depicts the analysis- and forecast-error changes as a function of the central location for the test added observations. Recall that all of the results shown are for changes in error averaged over the horizontal domain and at all levels, and that negative changes represent error reductions, i.e. improvements, while positive changes represent error increases, i.e. degradations. Figure 5 shows the error changes for a case with the same standard observation network (and thus the same ensemble spread) as in Fig. 3, but with a different realization of the initial-condition errors. Figures 6 and 7 show the standard observation network, the ensemble spread, and the error changes for the third sample case. Table 1 summarizes results for the 15 cases with this synoptic situation.

Figures 3–7 illustrate, in general, how assimilated added observations can influence analysis and forecast errors. Three examples are presented to emphasize the detail and complexity of the results. The examples show that small changes in the observation location can cause the results to vary significantly; in some situations, moving the added-observation cluster only a few hundred kilometres can change a much improved forecast into a much degraded forecast. The examples also show that changing the model initial conditions can cause the results to vary, even for the same synoptic situation. Comparing Fig. 4 with Fig. 7, we see that the influence of added observations, including the likelihoods, magnitudes, and specific patterns of improvements and degradations, depends on the past observation locations. Comparing Fig. 4 with Fig. 5, we see that the influence of adding observations also depends on the specific realization of the initial-condition errors. These aspects of the results, and their overall complexity, indicate why definitively diagnosing many of the degradations is difficult. They also show why one

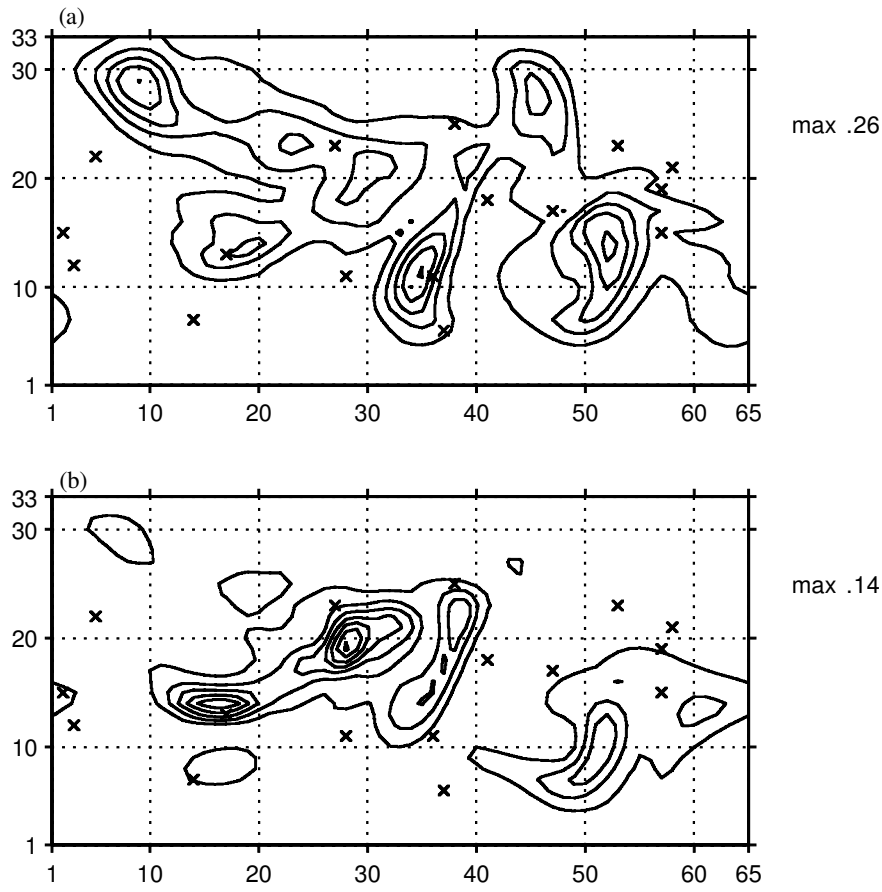

Figure 3. Ensemble spread in the 12-hour forecast valid at day 0 for standard observation network 1, at the uppermost interior level (a) and the lowest interior level (b), with an energy norm. The locations of the 16 fixed observations in the standard network are marked with an $\times$. The x-axis is longitude (periodic) and the y-axis is latitude. The ensemble has 97 members and is generated as described in section 2(c). The contour interval is 0.05 in (a) and 0.02 in (b), and the maximum values are given to the right. The tick mark labels on the axes represent grid points; the grid-point spacing is 250 km.

needs to test many sets of added observations before drawing firm conclusions about their effects.

Despite the differences, the cases have several features in common. First, all cases have mixed positive and negative results, with complex patterns of improvements and degradations and complex relationships between analysis-error changes and forecast-error changes. Second, most of the cases have more analysis improvements than degradations. As we see in Figs. 4, 5, and 7 and Table 1, the analysis improvements are also generally larger than the analysis degradations. Moving from analyses to 12-hour forecasts to longer term forecasts, however, the results tend to shift towards more degradations, and towards improvements and degradations being larger and of approximately equal magnitude. Finally, note that analyses and short-term forecasts are more likely to be improved when observations are added where the ensemble spread is large.

Table 2 summarizes results for the same case as in Fig. 4, but using different observation clusters for the added observations (see Morss 1999 for further detail). The results do depend to some extent on how the added observations are taken. For example,

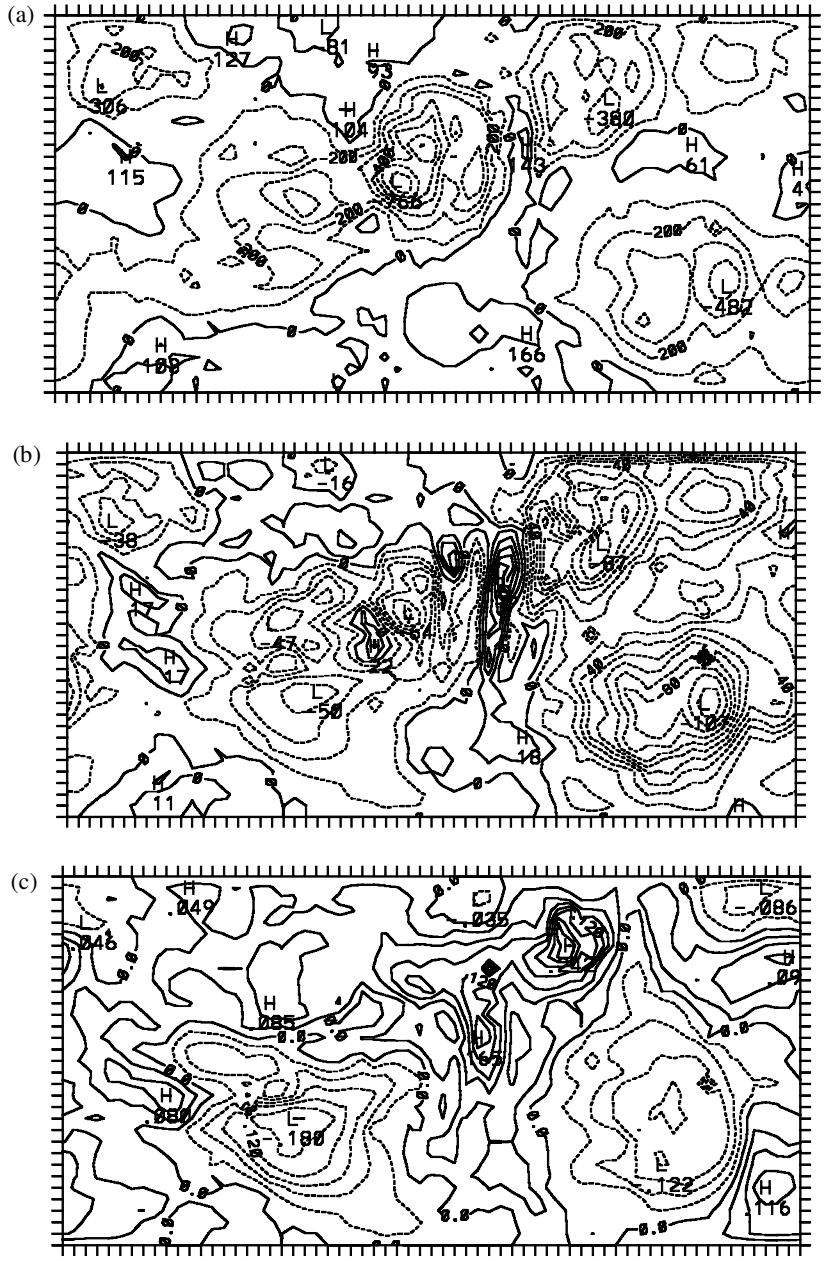

Figure 4. Analysis- and forecast-error change as a function of the central location for a test cluster of added observations in a sample case (day 0, standard observation network 1, initial realization 1), generated as described in section 3(b). Results are shown for the change in the global-average error in the analysis (a), 12-hour forecast (b), and 48-hour forecast (c); the x-axis is longitude (periodic) and the y-axis is latitude. Negative contour lines (improvements) are dashed; positive contour lines (degradations) are solid. The contour interval is 0.01 in (a) and (b) and 0.03 in (c), and the maximum and minimum values are summarized in Table 1. The synoptic situation is depicted in Fig. 1; the ensemble spread at the observation time and the standard observation network are depicted in Fig. 3.

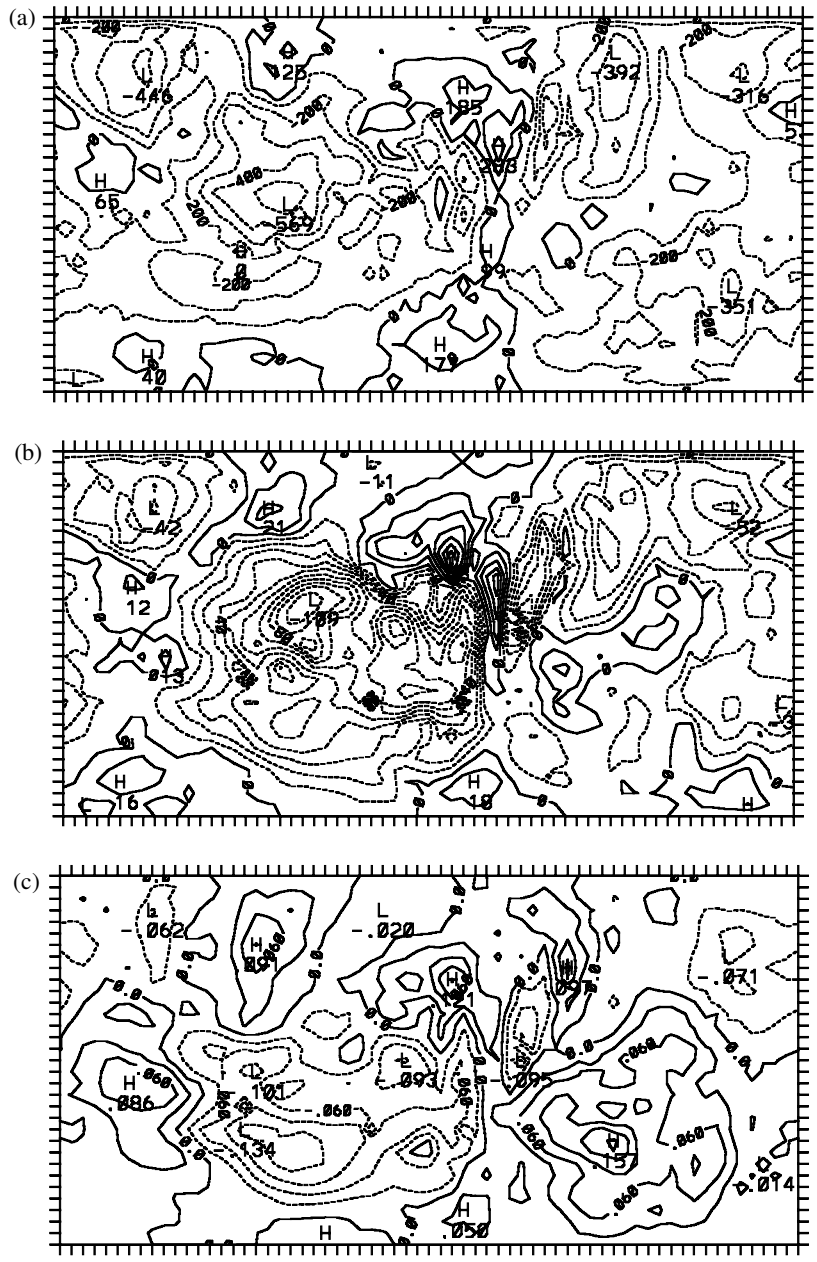

Figure 5. As in Fig. 4, for a different realization of the model state (initial realization 4).

the specific results at an observation location sometimes depend on the observation cluster used. In addition, clusters with more observations tend to produce not only larger improvements, but also larger degradations. Despite the differences, the general features discussed above are valid for all observation clusters tested.

Figures 4, 5, and 7 show how sensitive global-average analysis and forecast errors are to assimilated added observations. Adjoint models can be used to determine the

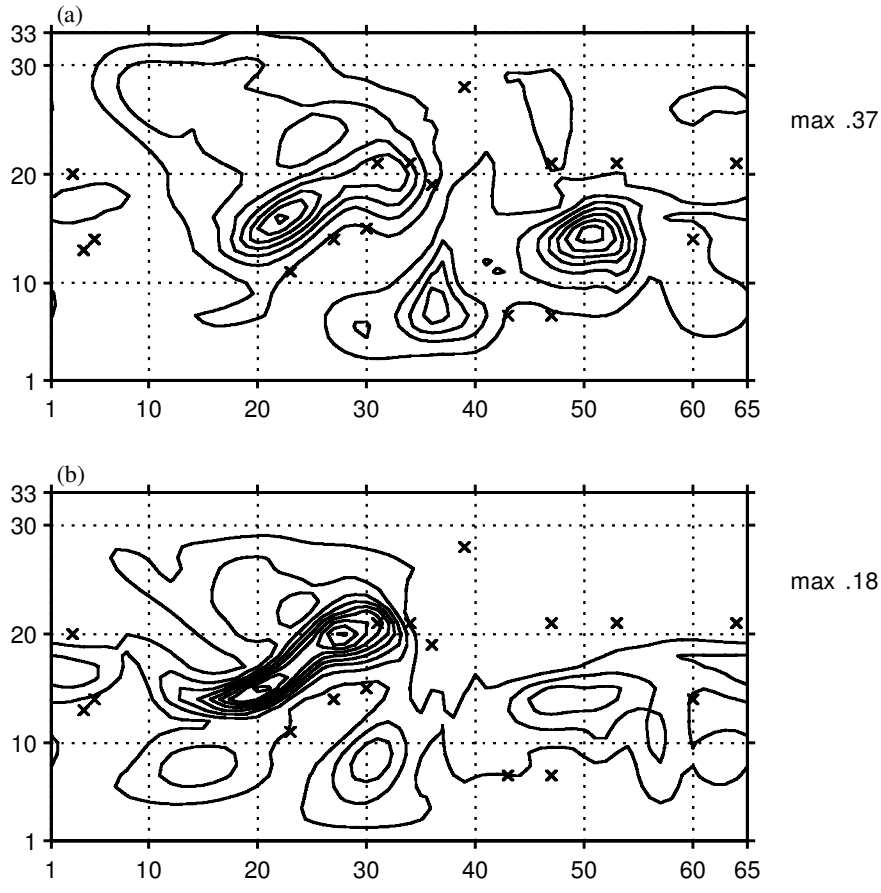

Figure 6. As in Fig. 3, for standard observation network 4.

sensitivity of forecast errors to small changes in the initial conditions (see, for example, Errico and Vukicevic 1992); the results in this study differ from adjoint-calculated sensitivities in several respects. First, in this study the full forecast model is used, while adjoints must approximate nonlinear parts of the forecast model. Unlike adjoint calculations, the results in this study also explicitly test the effects of the data-assimilation system. Because this includes extensive vertical and horizontal interpolation and extrapolation of the data, the results in this study tend not to have horizontal and vertical structure similar to that often seen in adjoint sensitivities, even when single observations are tested at single levels. Thus, although the results shown here and adjoint sensitivities are complementary and in some sense related, they are not directly comparable.

(*b*) *Results compiled over all cases*

Figure 8 shows, for all observation locations tested in the 255 cases, how frequently a set of observations changes the analysis and forecast errors by different amounts. Since the effect of adding observations is tested at each location in the domain in each case, a large fraction of the test observations are at locations that would probably not be selected by any targeting strategy, such as locations far from any current or future meteorological features. The results in Fig. 8, therefore, cannot be directly applied to any specific

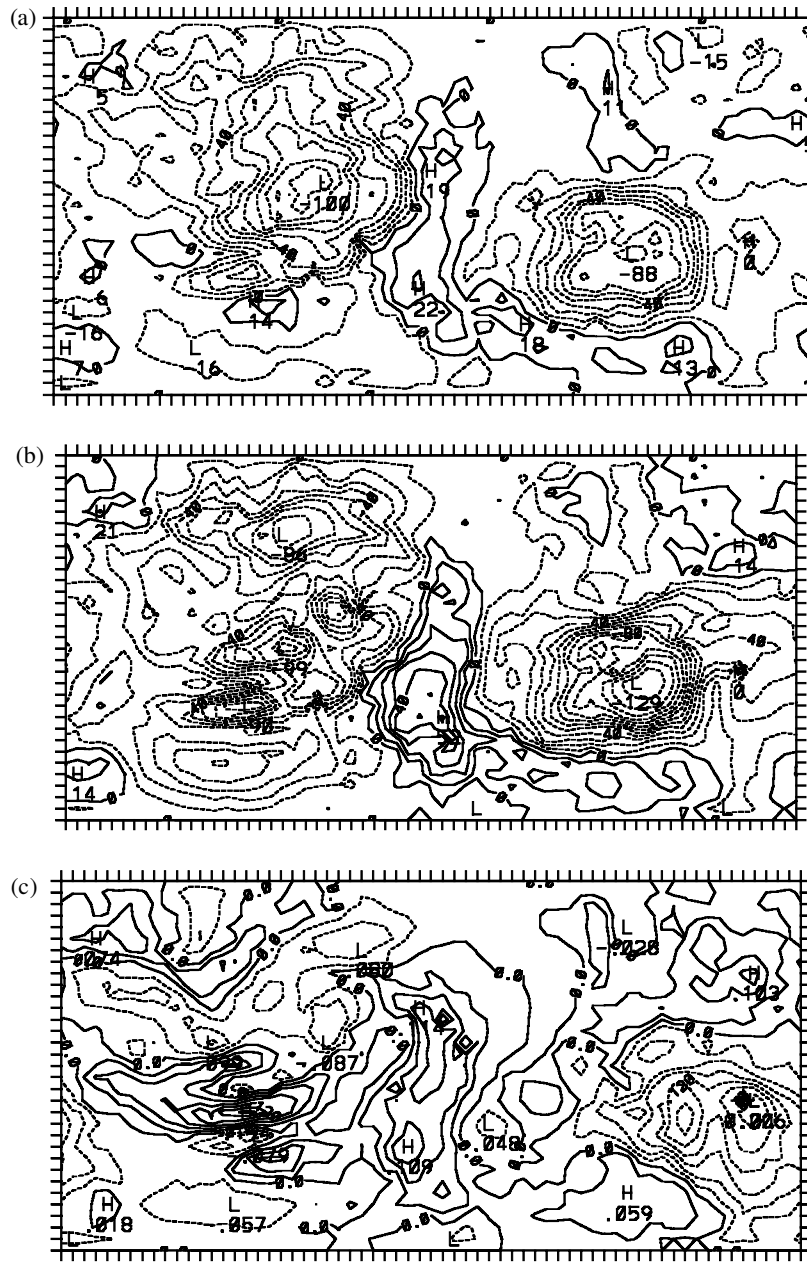

Figure 7. As in Fig. 4, for the standard observation network and ensemble spread depicted in Fig. 6.

targeted-observing strategy, but rather indicate the general likelihood of different sized improvements and degradations from randomly placed added observations.

Compiling the results over all cases confirms that several of the similarities among cases that were discussed in section 4(a) are general features. First, adding observations produces a mix of improvements and degradations. Adding observations also improves many more analyses than it degrades, and analysis improvements tend to be larger than

TABLE 1. MAXIMUM ERROR CHANGES, ADDED OBSERVATIONS AT DAY 0

Standard network	Initial realization	Analysis		12 h forecast		48 h forecast	
		Max. impr.	Max. degr.	Max. impr.	Max. degr.	Max. impr.	Max. degr.
1	1	−0.08	+0.02	−0.11	+0.06	−0.18	+0.24
1	2	−0.09	+0.04	−0.08	+0.08	−0.14	+0.16
1	3	−0.08	+0.02	−0.11	+0.06	−0.14	+0.12
1	4	−0.06	+0.03	−0.13	+0.06	−0.13	+0.16
1	5	−0.08	+0.03	−0.09	+0.06	−0.15	+0.24
1	6	−0.10	+0.03	−0.16	+0.05	−0.17	+0.18
1	7	−0.08	+0.03	−0.10	+0.08	−0.12	+0.17
1	8	−0.08	+0.01	−0.10	+0.05	−0.25	+0.14
1	9	−0.06	+0.03	−0.08	+0.08	−0.16	+0.10
1	10	−0.09	+0.01	−0.13	+0.04	−0.13	+0.23
2		−0.11	+0.03	−0.16	+0.06	−0.15	+0.13
3		−0.07	+0.05	−0.07	+0.07	−0.11	+0.19
4		−0.10	+0.02	−0.13	+0.05	−0.21	+0.19
5		−0.17	+0.03	−0.18	+0.04	−0.14	+0.19
6		−0.04	+0.07	−0.09	+0.11	−0.12	+0.16

Maximum improvements (reductions in global-average error) and degradations (increases in global-average error) produced by adding a cluster of observations to a network of 16 fixed observations at any location in the domain, in the synoptic situation shown in Fig. 1. Results are shown for changes in analysis errors, 12-hour forecast errors, and 48-hour forecast errors, for each of the 15 initial-condition errors tested. The change in error is averaged over all levels and is defined in Eq. (1). 'Standard network' refers to the network of fixed observations used to spin up the initial conditions, and 'initial realization' refers to the specific realization of the initial-condition errors, i.e. of the model state, as described in section 3.

TABLE 2. MAXIMUM ERROR CHANGES, DIFFERENT ADDED-OBSERVATION CLUSTERS, AT DAY 0

Cluster pattern	Analysis		12 h forecast		48 h forecast	
	Max. impr.	Max. degr.	Max. impr.	Max. degr.	Max. impr.	Max. degr.
1	−0.04	+0.03	−0.08	+0.06	−0.12	+0.08
3	−0.05	+0.02	−0.08	+0.05	−0.14	+0.13
13	−0.08	+0.02	−0.11	+0.06	−0.18	+0.24
81	−0.09	+0.02	−0.14	+0.09	−0.20	+0.36

As in Table 1, for standard observation network 1 and initial realization 1, but for observations added with different sized and shaped clusters. Cluster 13 is the cluster used for all other results shown, including those in Table 1; it is depicted in Fig. 2. Cluster 1 is a single rawinsonde at the test observation location, cluster 3 is a triangle of 3 rawinsondes around the test observation location, and cluster 81 is a rawinsonde at each of the 81 locations in a 9 grid point × 9 grid point (2000 km × 2000 km) area around the test observation location.

analysis degradations. Second, when forecasts are run from the analyses with added observations, both the magnitude of the error changes (positive or negative) and the likelihood of degradation tend to increase. Although degradations are more common in 12-hour forecasts than in analyses, few are large. By the 48-hour forecast, not only degradations, but also large degradations, are nearly as likely to occur as improvements of similar magnitude. In fact, as the forecast time-scale approaches the predictability time-scale, the distribution of forecast improvements and degradations must become symmetric; this is discussed further in section 6. Thus, although adding observations usually improves analyses and short-term forecasts, it often does not improve longer-term forecasts—even with a perfect forecast model.

As discussed in section 3(b), the results shown here are for error-free observations. For a subset of the cases, the experiments were also performed for test observations with errors*. Histograms similar to Fig. 8 for observations with errors show no significant

* Recall that no observations have errors greater than one standard deviation.

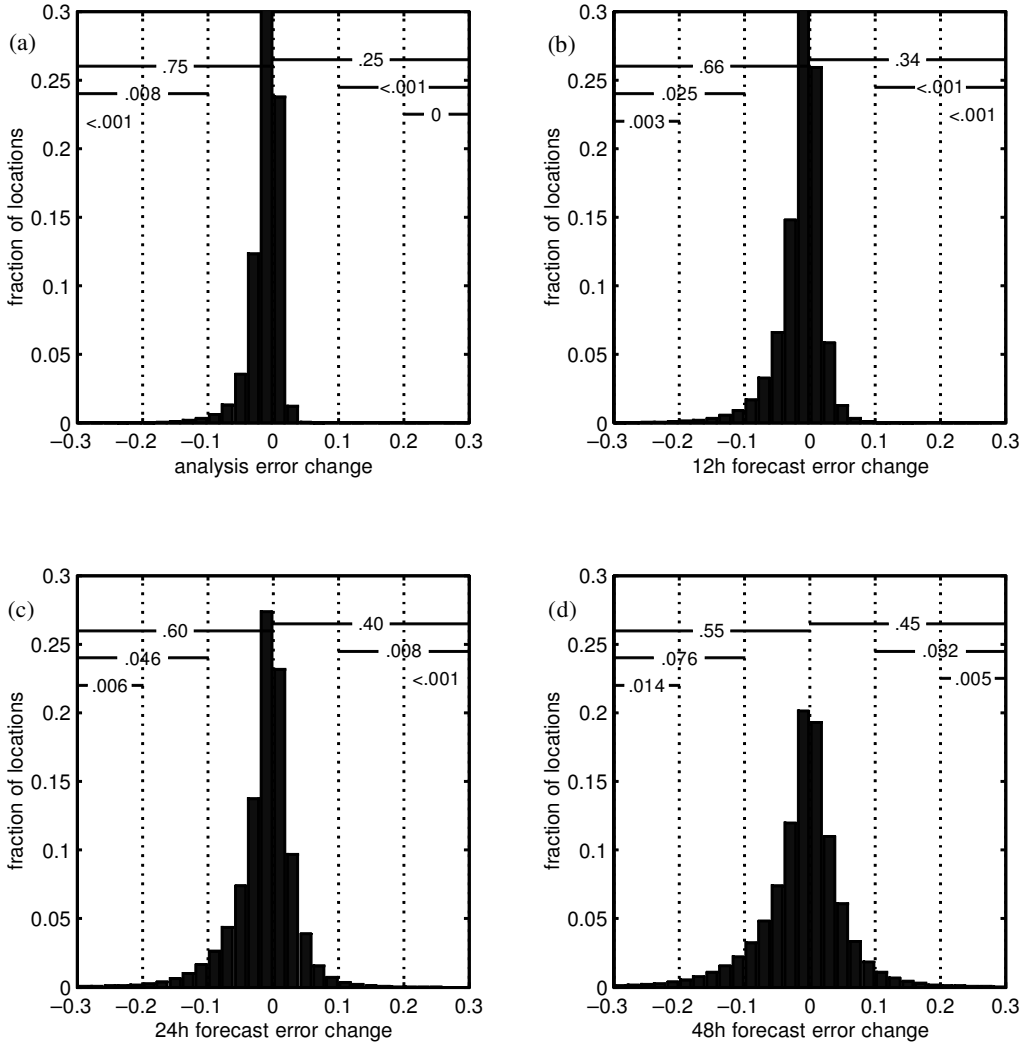

Figure 8. Histogram of the likelihood that a cluster of observations, added anywhere in the domain to a network of 16 fixed observations, changes the global-average error in the analysis (a), the 12-hour forecast (b), the 24-hour forecast (c) or the 48-hour forecast (d), by different fractions. Recall that negative changes represent improvements and positive changes represent degradations. The statistics are accumulated over 255 cases (15 initial-condition errors for each of 17 synoptic situations), as described in section 3. Observations are tested at 1968 locations in each case, for a total of $255 \times 1968 \approx 5 \times 10^5$ test sets of observations. The numbers near the top of the histogram are the fraction of locations at which added observations change the error by the amount in the indicated band. In (a), for example, at 75% of the locations, added observations reduced the analysis error by any amount; at 0.8% of the locations, added observations reduced the analysis error by more than 10%. A 0 means that there are no locations in that section of the histogram; <0.001 means that there are a few locations, but less than 0.1% of the number tested.

differences from those for observations without errors (compiled over the same subset of cases). Compared with observations without errors, observations with errors cause approximately 4% more analysis degradations, and approximately the same percentage of degradations in forecasts of one day or longer. Thus, including errors in the observations does not significantly change either the size distribution of error changes in analyses or forecasts or the likelihood of analysis or forecast degradations. This suggests that, at

least for the observation and background errors simulated here, observation errors are only a minor factor in analysis and forecast changes. As the average background error decreases relative to the average observation error, perhaps due to a denser standard observation network or a better data-assimilation system, observation errors are expected to play a larger role. The role of observation errors in degradations is discussed further in section 5.

(*c*) *Expected results of adding observations*

The results shown so far depict how adding observations affects individual realizations of analyses and forecasts. In practice, however, one can only predict expected quantities. Therefore, in this section we briefly discuss results similar to those in sections 4(a) and 4(b), but for the expected change in error from adding observations in a case. To estimate the expected change in error, at each test location we average the results from several equally likely realizations of the initial-condition errors for the same synoptic situation and standard observation network*. This represents the results one would expect if added observations were assimilated into several ensemble members, forecasts were produced for each ensemble member with and without the added data, and the comparisons were averaged over the ensemble.

Figure 9 shows the expected effect of adding observations for the same synoptic situation, standard observation network, and ensemble spread as the results in Figs. 4 and 5. Note that the expected results tend to have smaller magnitudes than the results for individual realizations. This indicates that several of the features in Figs. 4 and 5 are due to the specific realization of the initial-condition errors. In addition, as the forecast length increases, so does the difference in magnitude between the expected results and those in individual realizations. As discussed in sections 4(b) and 6, this indicates that error growth tends to weaken the relationship between changes in the initial conditions and changes in longer-term forecasts.

Despite the differences among results for different realizations, several features are evident in the expected results. First, the expected results identify a region near the centre of the domain where adding observations, on average, significantly degrades the 12-hour forecast (and where most realizations have 12-hour forecast degradations). Although ten realizations may be insufficient to allow one to draw firm conclusions, the mean degradation is probably caused by a mismatch between the true and assumed background-error statistics in this region, combined with the error growth characteristics of the synoptic situation. Second, as in the results for individual cases, analysis and 12-hour forecast improvements are more likely and tend to be larger when observations are added where the ensemble spread is larger.

We also constructed histograms similar to Fig. 8 for the expected results (not shown). Compared with histograms for the results from individual realizations (compiled over the same subset of cases), the expected results contain approximately 12% fewer analysis and 12-hour forecast degradations. For 48-hour forecasts, the expected results contain only several percent fewer degradations than the results for individual realizations, but are clearly more skewed towards large improvements. Thus, measuring the effect of adding observations averaged over several realizations of an analysis and forecast decreases the likelihood of degradation. It also tends to decrease the magnitude of degradations more than it decreases the magnitude of improvements. This suggests that adding

* The results shown are for the error norm defined in Eq. (1), the percent change in error. The percent change is defined using the error without added observations, which is different for each realization. The expected change in error was also calculated by averaging over the absolute changes in error for the realizations; the results are similar to those shown.

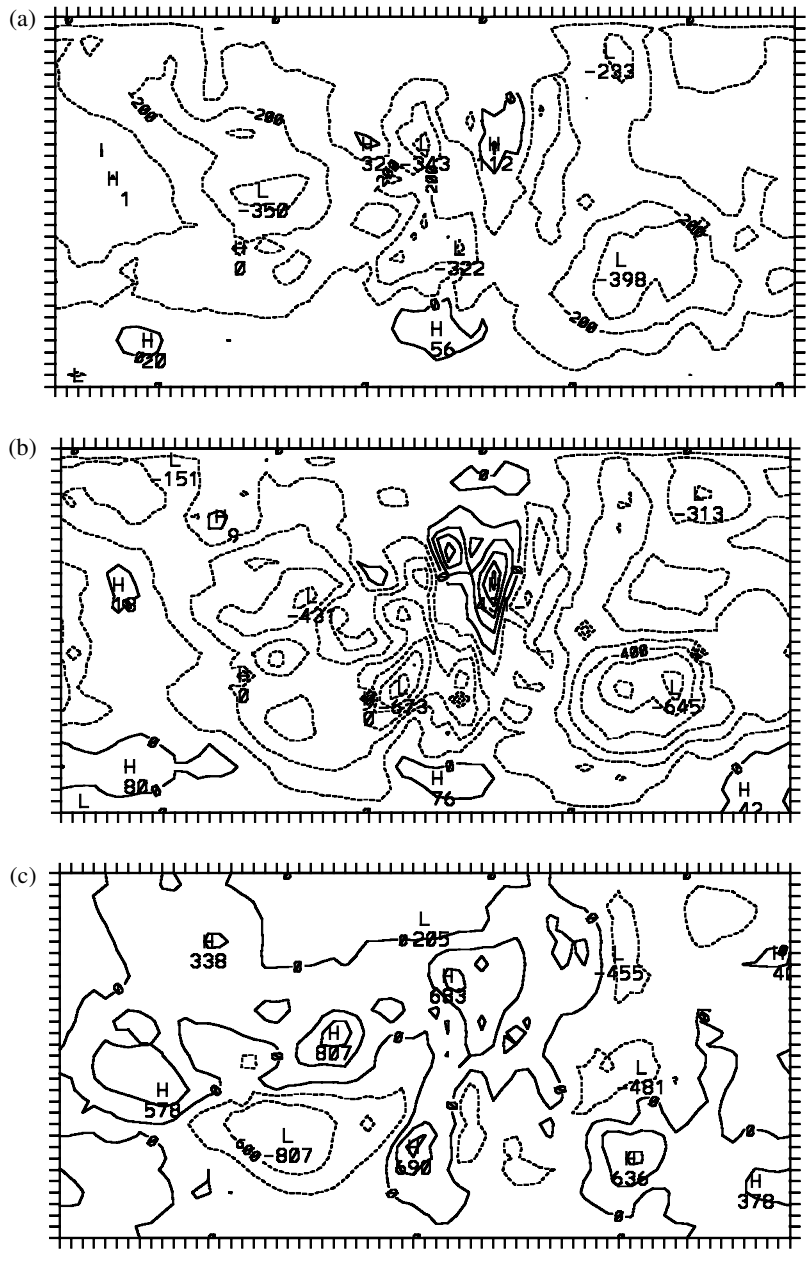

Figure 9. As in Figs. 4 and 5, but for the expected change in error with added observations. The expected change in error is calculated by averaging over the results for the ten equally likely realizations of the initial-condition errors, in other words, by averaging the results in Figs. 4 and 5 with the results from the eight other realizations of initial-condition errors tested for the synoptic situation and standard observation network depicted in Figs. 1 and 3. Figures 4 and 5 suggest the variability about the expected results. The contour intervals are the same as in Fig. 4.

or targeting observations may benefit probabilistic forecasts more than deterministic forecasts.

5. Analysis degradations in idealized low-dimensional data-assimilation systems

The results shown so far use a data-assimilation system with error statistics that are far from accurate. In this section, we begin to evaluate how adding observations can affect analyses in other data-assimilation systems, including those with much better error statistics. To do so, we construct statistical data-assimilation systems that are similar to the 3D-Var or a Kalman filter but much simpler, with only one or two degrees of freedom. We then use the low-dimensional data-assimilation systems to explore how, in general, data-assimilation systems incorporate information in different circumstances. Testing simple, time-independent data-assimilation systems allows us to avoid many of the complex interactions in the 3D-Var scheme, and it greatly facilitates simulating perfect knowledge of the error statistics.

Each low-dimensional data-assimilation experiment has at least two input variables, one representing a background estimate and one representing an added observation. The input variables are random variables, not designed to represent a physical quantity. To test each data assimilation over many possible situations, first, 10^6 realizations are generated for each input variable. Second, for each of the 10^6 sets of input variables, an analysis is produced. Then, for each set of realizations, the estimates with and without the added observation(s) are evaluated and compared.

The assimilation algorithms tested here, like 3D-Var or a Kalman filter, assume that the errors in the input variables are random with a normal (Gaussian) distribution. Thus, the realizations for a variable x are generated by randomly sampling from a Gaussian distribution with mean x^t and variance σ^2; this is written $x = x^t + \epsilon$, $\epsilon \sim N(0, \sigma^2)$. Variables that are correlated are generated jointly using the covariance matrix.

Although one could easily experiment with many idealized low-dimensional data-assimilation algorithms, results are discussed for only a few examples. All of the results discussed simulate perfect knowledge of the observation-error statistics, in other words generate the realizations of the observation input variables using the error distributions assumed by the data-assimilation algorithm. The first two sections also simulate perfect knowledge of the background-error statistics; in the first, we estimate one analysis variable at an observed location, and in the second, we extrapolate information from an observed to a unobserved location. The third section explores some effects of imperfect knowledge of the background-error statistics.

The low-dimensional data-assimilation examples identify three factors that can cause analysis degradations: observation errors, extrapolation of information, and imperfect error statistics. The first two factors can cause degradations even with perfect knowledge of the error statistics; thus, the risk of degradation in individual realizations of analyses is inherent in statistical data assimilation. The examples also show that current statistical data-assimilation systems are designed to minimize the mean error in the estimate and the standard deviation of the error in the estimate, but not the likelihood of degradation. In addition, they identify several general circumstances in which adding observations is more likely to improve an estimate and on average improves an estimate more.

Although the results in this section are not directly comparable to those from high-dimensional data-assimilation systems, they are shown for several reasons. First, they

confirm that the degradations shown so far are not merely a feature of the specific data-assimilation system used. Second, they demonstrate several general principles of how any realistic statistical data-assimilation system is likely to use information in different circumstances*. These principles are used in section 7 to help us interpret results from the 3D-Var model.

(*a*) *One analysis variable, one observation: Perfect error statistics*

Suppose that at a single location and time, we have two unbiased estimates of a variable that has a true value x^{t}. One estimate, the random variable $x^{\mathrm{b}} = x^{\mathrm{t}} + \epsilon^{\mathrm{b}}$, $\epsilon^{\mathrm{b}} \sim N(0, \sigma_{\mathrm{b}}^2)$, represents the background value, the estimate of x^{t} in the absence of further information. x^{b} could be, for example, a climatological value, a forecast from a previous time, or an analysis with many observations. The second estimate, $x^{\mathrm{o}} = x^{\mathrm{t}} + \epsilon^{\mathrm{o}}$, $\epsilon^{\mathrm{o}} \sim N(0, \sigma_{\mathrm{o}}^2)$, represents a new observation of x^{t}.

If we know the statistics of x^{b} and x^{o} correctly, then the minimum variance and maximum likelihood estimate of x^{t} is:

$$x^{\mathrm{a}} = \frac{\sigma_{\mathrm{b}}^2 x^{\mathrm{o}} + \sigma_{\mathrm{o}}^2 x^{\mathrm{b}}}{\sigma_{\mathrm{b}}^2 + \sigma_{\mathrm{o}}^2} = x^{\mathrm{b}} + \frac{\sigma_{\mathrm{b}}^2}{\sigma_{\mathrm{b}}^2 + \sigma_{\mathrm{o}}^2}(x^{\mathrm{o}} - x^{\mathrm{b}}). \tag{2}$$

x^{a} represents the analysis, our 'best' estimate of x^{t} given a realization of x^{b} and x^{o}. $x^{\mathrm{a}} - x^{\mathrm{b}}$ (the last term on the right-hand side of Eq. (2)) is the analysis increment. For a derivation of Eq. (2) or the related multivariate algorithm used in 3D-Var, a Kalman filter, or other statistical data-assimilation systems, see, for example, Lorenc (1986) or Daley (1991).

For each realization, define the error in x^{b}, i.e. in the estimate without the added observation x^{o}, as

$$e^{\mathrm{b}} = |x^{\mathrm{b}} - x^{\mathrm{t}}| \tag{3}$$

and the error in x^{a}, i.e. in the estimate with the added observation, as

$$e^{\mathrm{a}} = |x^{\mathrm{a}} - x^{\mathrm{t}}|. \tag{4}$$

Then, define the change in error produced by assimilating the added observation as

$$\Delta e = e^{\mathrm{a}} - e^{\mathrm{b}}. \tag{5}$$

Equation (5) is similar to Eq. (1), but for absolute change in error rather than percent change. Again, $\Delta e < 0$ when assimilating the added observation improves the analysis, and $\Delta e > 0$ when assimilating it degrades the analysis.

Figure 10 shows the distribution of Δe for 10^6 realizations of this data-assimilation example, when $\sigma_{\mathrm{b}} = \sigma_{\mathrm{o}} = 1$. As expected, assimilating the observation *on average* improves the estimate: the mean of e^{a} is 0.56, while the mean of e^{b} is 0.80. Assimilating the observation also decreases the uncertainty in the estimate:

$$\sigma_{\mathrm{a}}^2 = \left(\frac{1}{\sigma_{\mathrm{b}}^2} + \frac{1}{\sigma_{\mathrm{o}}^2}\right)^{-1} = 0.5 \tag{6}$$

* One could construct a data-assimilation example with highly non-Gaussian error statistics that might behave differently; such systems are not considered here.

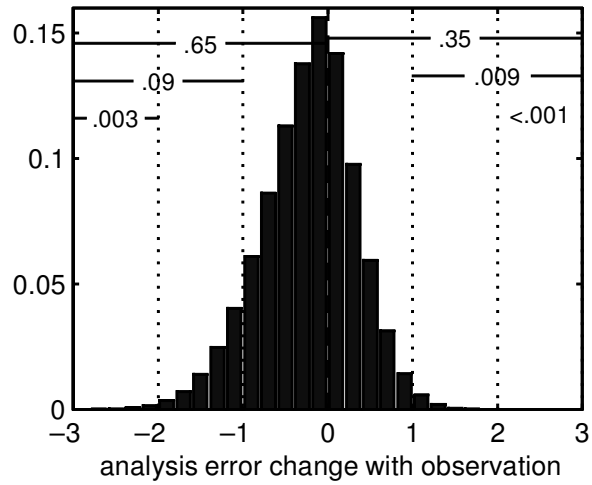

Figure 10. Histogram of the likelihood that assimilating one observation into an estimate of the observed variable changes the analysis error by different amounts. The numbers near the top are the fraction in the indicated band. See text for further details.

is less than $\sigma_b^2 = 1$. However, assimilating the observation does not *always* improve the estimate: adding an observation degrades the analysis, that is, e^a is larger than e^b, approximately 35% of the time. Recall that the error statistics used in Eq. (2) to calculate x^a match the actual statistics of x^b and x^o, so none of the degradations are caused by incorrect specification of the error statistics in the data-assimilation algorithm. Rather, the degradations occur because the assimilation is statistical—having perfect information about the statistics of the errors is not the same as having perfect information to correct the errors.

Degradations occur when $e^a > e^b$, i.e. when

$$|x^a - x^t| > |x^b - x^t|. \tag{7}$$

To visualize how the degradations occur, we combine Eq. (7) with Eq. (2), then solve for x^o. For $x^b > x^t$, assimilating the observation degrades the estimate in two situations:

$$x^o > x^b \tag{8}$$

$$x^o - x^t < -\{1 + 2(\sigma_o^2/\sigma_b^2)\}(x^b - x^t). \tag{9}$$

The two situations are depicted in Fig. 11; the equivalent situations for $x^b < x^t$ can be visualized by flipping Fig. 11 left to right. Note that in this example, which estimates one observed variable, incorporating an error-free observation cannot degrade the estimate.

Figure 12 depicts how the likelihood of degradation and the mean change in error in this data-assimilation example vary with the ratio of σ_b to σ_o. The results show that (with perfect knowledge of the error statistics) improvements are more likely and the mean change is a larger improvement when observations are added where the ratio between the background error and the observation error is likely to be large. As σ_b/σ_o decreases, the situations in Eqs. (8) and (9) become more likely, increasing the likelihood of degradation; to visualize how this occurs, picture the distribution of x^b narrowing compared with that of x^o in Fig. 11. When σ_b/σ_o is small, observations are nearly as likely to degrade analyses as to improve them, and the mean change is only a small improvement. This makes sense intuitively—as the background error decreases relative

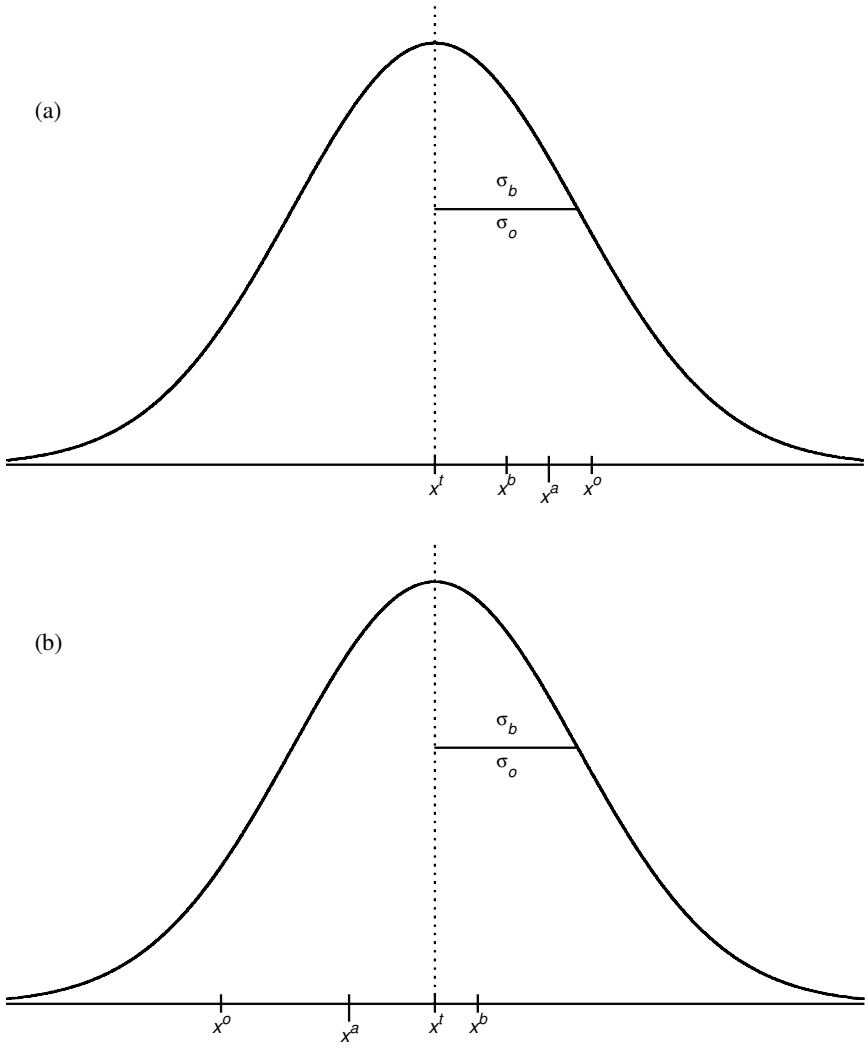

Figure 11. Illustration of how adding an observation can degrade an estimate of one variable for some realizations of x^{b} and x^{o}, even when the data-assimilation system has perfect knowledge of the error statistics. (a) Illustrates a degradation in the situation given by Eq. (8), and (b) illustrates a degradation in the situation given by Eq. (9). The examples shown are for $\sigma_{\mathrm{b}} = \sigma_{\mathrm{o}}$. See text for further details.

to the observation error, the noise in the observation increases relative to the error the observation is trying to correct.

Figure 12 shows that, when observing in regions with small background errors, we will require good observations to have a good chance of improving the analysis. To increase the chance of improving the analysis, the observation error could be decreased, by reducing either the instrument error or the representativeness error. Or, if the observation error is fixed, several measurements could be taken. To simulate taking several measurements, we have rerun the same set of experiments, but assimilating more than one independent observation x^{o} of the same type (same variable, same σ_{o}, same location). As the number of observations increases, the uncertainty in the estimate, the mean error in the estimate, and the likelihood of degradation all decrease. Even when

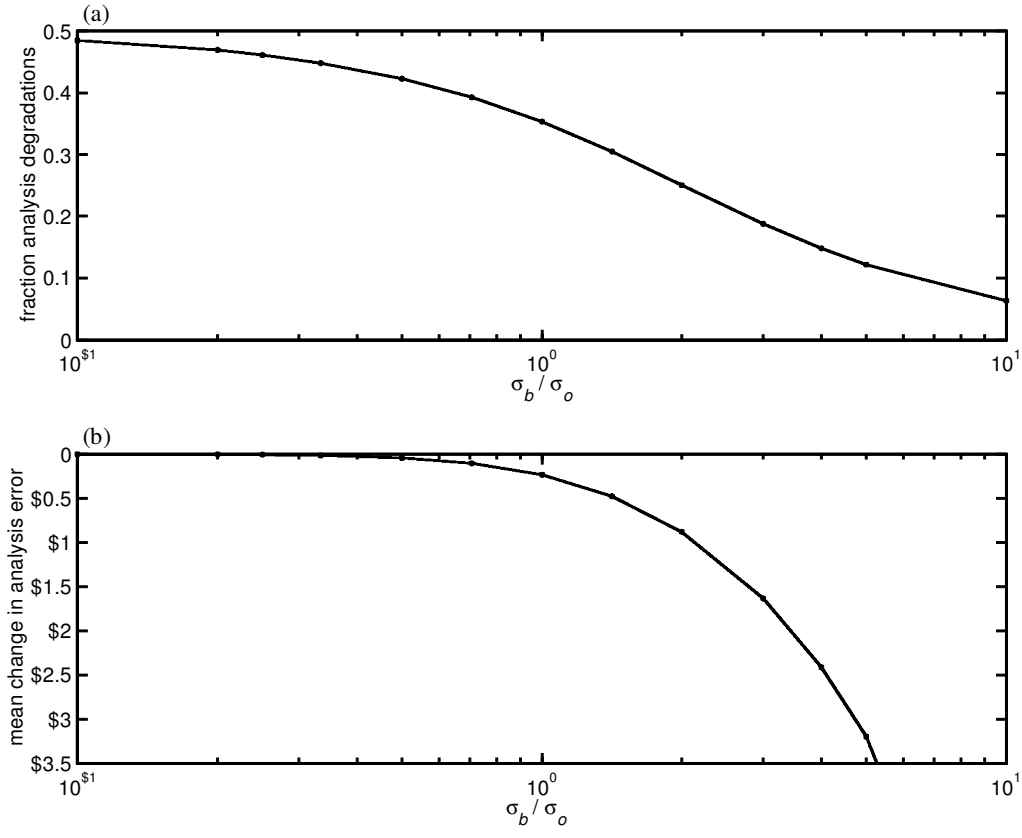

Figure 12. Fraction of changes that are degradations (a) and mean change in analysis error (mean of Δe, (b)) as a function of the ratio between σ_b (the standard deviation of the background error) and σ_o (the standard deviation of the observation error) in the one-variable data-assimilation example given by Eq. (2), with perfect knowledge of the error statistics. Each point plotted is an average over 10^6 realizations. (b) Is normalized to $\sigma_o = 1$. Note that the x-axis is logarithmic.

many independent observations are assimilated, however, degradations are still possible; with $\sigma_o = \sigma_b$ and 100 observations, for example, approximately 6% of the changes are degradations.

The results from the one-variable data assimilation have several implications for any statistical data-assimilation system. First, even with perfect knowledge of all characteristics of the error statistics, adding an observation can degrade an individual realization of an analysis. Second, observations are more likely to improve analyses, and they on average improve analyses more, when the observation error is likely to be small compared with the background error. Again, this is true even if the data-assimilation system uses the correct error statistics. Finally, degradations in single realizations are difficult to avoid completely, even with many observations.

(*b*) *Two analysis variables, one observation: Perfect error statistics*

In the atmosphere, variables tend to be spatially correlated. Estimation of the atmospheric state is also generally thought to be an underdetermined problem, in other words, to have fewer observations than degrees of freedom. In this section, we extend the data assimilation of section 5(a) to simulate these two features by assimilating one

observation at both the observed location and an unobserved location. The results can also represent extrapolating information from one observation to any analysis variable whose errors are on average correlated with, but not the same as, the errors in the analysis of the observed variable.

Suppose we have one observation, $x_1^{\mathrm{o}} = x_1^{\mathrm{t}} + \epsilon_1^{\mathrm{o}}$, $\epsilon_1^{\mathrm{o}} \sim N(0, \sigma_{\mathrm{o}1}^2)$, and two background variables: x_1^{b} at the observation location, and x_2^{b} at another location (where the true value is x_2^{t}). x_1^{b} and x_2^{b} are unbiased and have errors that are normally distributed with standard deviations of $\sigma_{\mathrm{b}1}$ and $\sigma_{\mathrm{b}2}$, respectively. The mean correlation between the errors in x_1^{b} and x_2^{b} is given by c_{12}, where $-1 \leq c_{12} \leq 1$.

The analysis at location 1 is calculated using Eq. (2), and the results are the same as those in section 5(a). The analysis at location 2 is calculated using the minimum variance and maximum likelihood estimate of x_2^{t}:

$$x_2^{\mathrm{a}} = x_2^{\mathrm{b}} + \frac{\sigma_{12}^2}{\sigma_{\mathrm{b}1}^2 + \sigma_{\mathrm{o}1}^2}(x_1^{\mathrm{o}} - x_1^{\mathrm{b}}), \tag{10}$$

where

$$\sigma_{12}^2 = c_{12}\sigma_{\mathrm{b}1}\sigma_{\mathrm{b}2}. \tag{11}$$

This is a simplification of the 3D-Var or a Kalman filter to one observation and two analysis variables. Again, the analysis is performed for 10^6 realizations of the input variables, and the realizations are generated using the same error statistics as those assumed in the data-assimilation algorithm. The errors are defined as in section 5(a).

Figure 13 shows the results of the analysis at the unobserved location for different values of $|c_{12}|$, when $\sigma_{\mathrm{o}1} = \sigma_{\mathrm{b}1} = \sigma_{\mathrm{b}2}$*. When $|c_{12}| = 1$, the errors in x_1^{b} and x_2^{b} are perfectly correlated, and the results are the same as those in section 5(a). As $|c_{12}|$ decreases, both the likelihood of improvement and the mean improvement decrease. When $|c_{12}|$ is small, degradations are nearly as likely as improvements; however, because the magnitude of the analysis increment is proportional to $|c_{12}|$, these degradations tend to be small. In the limit $c_{12} = 0$, $x_2^{\mathrm{a}} = x_2^{\mathrm{b}}$; there are no improvements or degradations.

The circumstances in which x_2^{a} has a larger error than x_2^{b} can be worked out analytically and depicted, as in section 5(a). Here, we note only that the estimate at location 2 can be degraded in two general circumstances: when the estimate at location 1 is degraded, or, if $|c_{12}| \neq 1$, when the estimate at location 1 is improved. The first can occur only when the observations include errors; the second can occur for error-free observations. Unlike at an observed location, therefore, assimilating an error-free observation at an unobserved location can degrade the analysis—even with perfect knowledge of the observation- and background-error statistics.

These results show that even a 'perfect' statistical data-assimilation system with very good observations can degrade some aspects of an analysis if it has degrees of freedom that are unobserved. The results also indicate that when the average correlation between background errors at an observed location and those at an unobserved location is smaller, extrapolating data to the unobserved location carries a greater risk of degradation. This is likely to be important in more realistic, high-dimensional forecasting systems, when data-assimilation procedures must reconstruct a complex and rapidly changing atmospheric state from limited information.

* When $\sigma_{\mathrm{o}1} = \sigma_{\mathrm{b}1} \neq \sigma_{\mathrm{b}2}$, Fig. 13 is nearly identical; panel (a) is the same, and panel (b) has the same shape, with different values on the y-axis. When $\sigma_{\mathrm{o}1} \neq \sigma_{\mathrm{b}1} = \sigma_{\mathrm{b}2}$, Fig. 13 looks similar, but as $|c_{12}|$ approaches 1, the likelihood of degradation and the mean change in error approach the values for the one-variable example with the appropriate ratio of $\sigma_{\mathrm{b}1} : \sigma_{\mathrm{o}1}$.

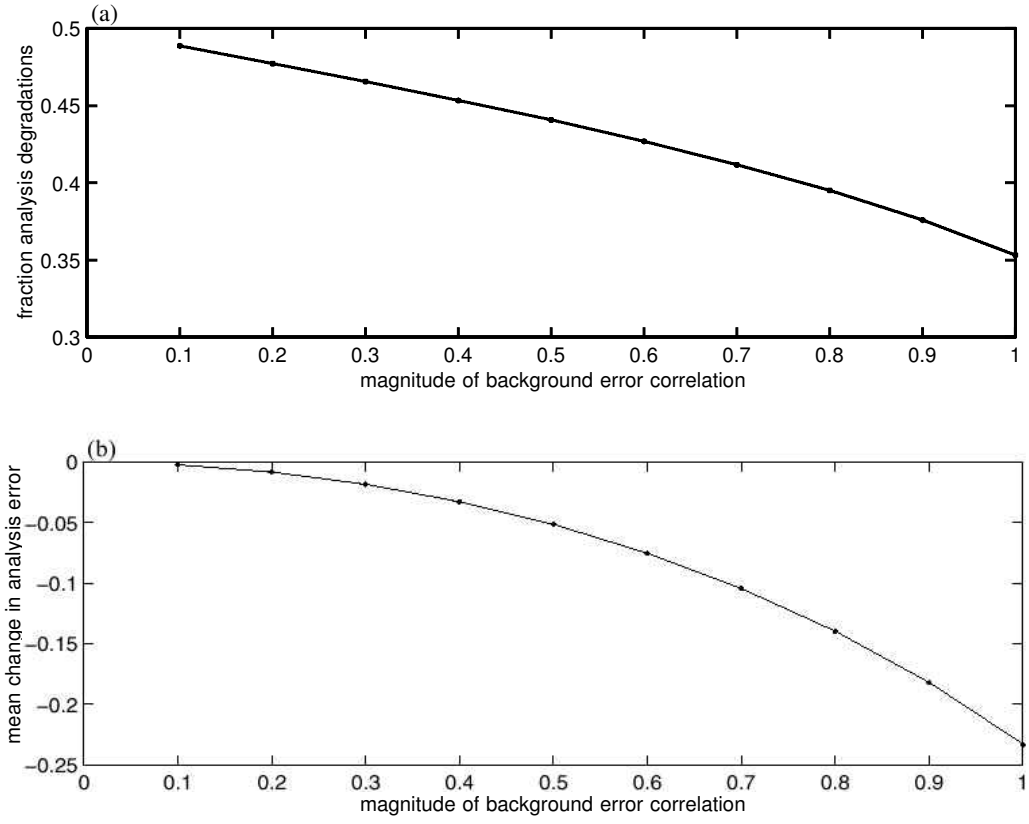

Figure 13. As in Fig. 12, but for the results of assimilating an observation at an unobserved location as a function of $|c_{12}|$, the magnitude of the mean correlation between the background errors at the observed location and at the unobserved location. The assimilation uses the algorithm given by Eq. (10), with perfect knowledge of the error statistics. Results are shown for $\sigma_{b1} = \sigma_{b2} = \sigma_o$; (b) is normalized to $\sigma_o = 1$.

(*c*) *Imperfect background-error statistics*

Next, we briefly explore several effects of imperfect knowledge of the background-error statistics. To do so, we rerun the examples from sections 5(a) and 5(b), but specify the background-error statistics incorrectly in the data-assimilation algorithm. In the results discussed here, all of the error distributions remain Gaussian; in other words, we simulate knowing the shapes of the distributions but not the variances and covariances. Again, each example is tested over 10^6 realizations of the input variables, and errors are defined as in section 5(a).

For the one-variable data assimilation from section 5(a), we mis-specify the background-error statistics by using the distribution $N(0, \sigma_b^2)$ to generate x^b, but an estimate of σ_b, σ_b^{est}, in Eq. (2) to calculate x^a. Figure 14 shows the results for different ratios of σ_b^{est} to σ_b, when $\sigma_o = \sigma_b$. As expected, the mean improvement in the estimate is maximized and the variance in the estimate (not shown) is minimized when $\sigma_b^{est} = \sigma_b$. The likelihood of degradation, however, is minimized not when $\sigma_b^{est} = \sigma_b$, but when $\sigma_b^{est} \ll \sigma_b$.

Why does the likelihood of degradation increase as σ_b^{est}/σ_b increases? Recall from section 5(a) that assimilating an observation can degrade an estimate of the observed

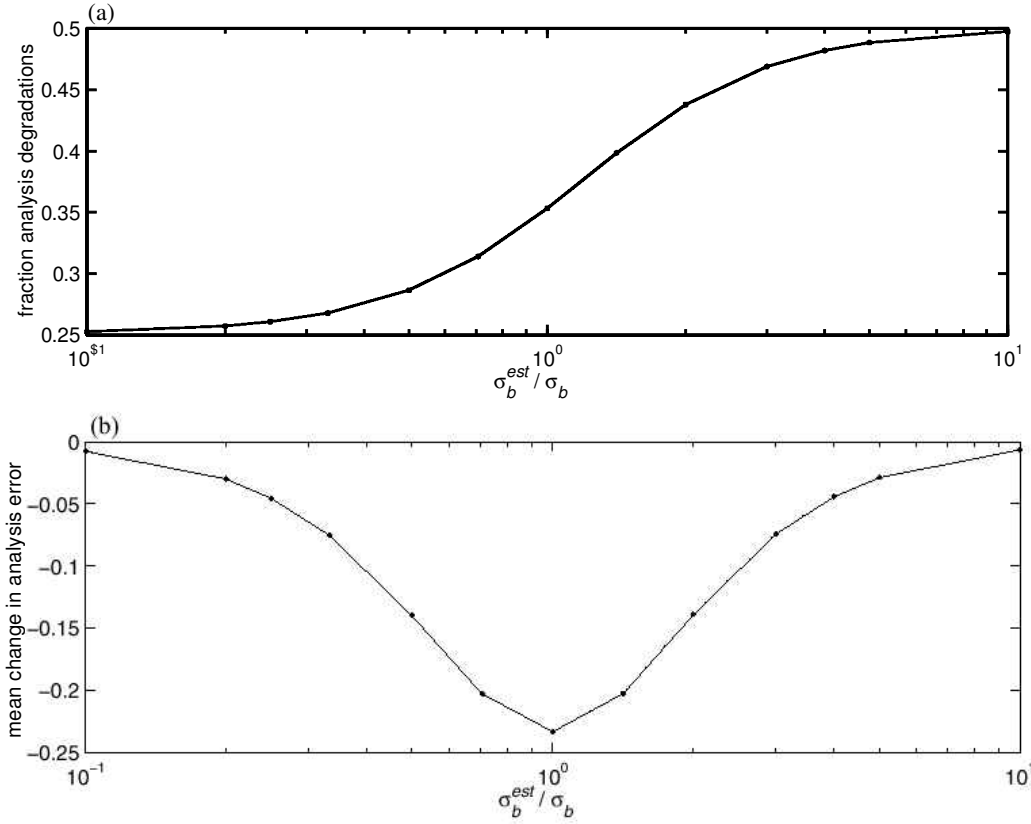

Figure 14. As in Fig. 12, for results as a function of the ratio between σ_b^{est} and σ_b in the one-variable data assimilation with imperfect knowledge of the error statistics. σ_b is the actual standard deviation of the background error, used to generate the realizations of x^b, and σ_b^{est} is the estimated standard deviation of the background error, used in Eq. (2) to calculate x^a. Results are shown for $\sigma_b = \sigma_o$; (b) is normalized to $\sigma_o = 1$.

variable in two ways. Substituting σ_b^{est} for σ_b in Eqs. (8) and (9), we see that, for a given distribution of x^b and x^o, the likelihood of the first type of degradation is independent of σ_b^{est}. The second type of degradation, on the other hand, becomes more likely as σ_b^{est} increases. To visualize why the likelihood of degradation increases, note that increasing σ_b^{est} relative to σ_b increases the magnitude of the analysis increment, drawing the analysis closer to the observation; in Fig. 11, this does not affect the first type of degradation, but increases the likelihood of the second type of degradation. Thus, as σ_b^{est}/σ_b increases, degradations become more likely.

When $\sigma_b^{est} \gg \sigma_b$, the assimilation, thinking that the background errors are large, overweights the observation. This produces analysis increments that are much larger than the average background error, so both improvements and degradations tend to be large. Because degradations are nearly as likely as improvements, however, the improvements and degradations tend to cancel, leaving a small mean change in error. When $\sigma_b^{est} \ll \sigma_b$, on the other hand, the assimilation thinks that background errors are small; it overweights the background variable, producing small analysis increments. In this situation, improvements are frequent, but they tend to be small.

These results suggest that one can decrease the likelihood that an added observation will degrade a single realization of an analysis by underestimating the background

error. Unfortunately, this also decreases the average magnitude of the improvements. Overestimating the background error produces larger improvements, but it also produces more frequent and larger degradations. Thus, when estimating σ_b to assimilate added data at one time to produce a single realization, there may be a trade-off between decreasing the likelihood of analysis degradation and increasing the potential for a large analysis improvement*.

For the two-variable data assimilation in section 5(b), we first simulate knowing σ_{b1} and σ_{b2}, but using an estimate of c_{12}, c_{12}^{est}, in Eqs. (10) and (11) to calculate x_2^a. For simplicity, we discuss results for $\sigma_{o1} = \sigma_{b1} = \sigma_{b2}$. When c_{12}^{est} and c_{12} have the same sign, the general results are similar to those in Fig. 14, with $|c_{12}^{est}|/|c_{12}|$ on the x-axis; the mean change in error at the unobserved location is minimized when $c_{12}^{est} = c_{12}$, but the likelihood of degradation is minimized when $|c_{12}^{est}| \ll |c_{12}|$. Overestimating $|c_{12}|$ increases the risk of degradation because it increases the magnitude of the analysis increment, as overestimating σ_b does in the one-variable example. One can overestimate $|c_{12}|$ by, for example, overestimating the correlation length of background errors.

In some circumstances, mis-specifying c_{12} can cause assimilating an added observation to degrade more estimates than it improves and to, on average, degrade the estimate. When c_{12}^{est} and c_{12} have the same sign, this occurs when $|c_{12}| \leqslant 0.5$ and $|c_{12}^{est}| > \sim 2|c_{12}|$. When c_{12}^{est} and c_{12} have opposite signs, this occurs for all values of c_{12}^{est} and c_{12}. Thus, if a background-error correlation is significantly overestimated or is estimated with the wrong sign, at least some aspects of the analysis will, on average, be more accurate if an added observation is not assimilated.

The two-variable data-assimilation example is also affected by estimates of σ_{b1} or σ_{b2}. Combining Eq. (10) with Eq. (11), we see that σ_{b1} appears in both the numerator and the denominator of the analysis increment at location 2, while σ_{b2} appears only in the numerator. Estimating σ_{b1} incorrectly thus affects the analysis increment at the unobserved location only indirectly, and has an indirect effect on the results. Estimating σ_{b2} incorrectly, on the other hand, affects the analysis increment directly. Overestimating σ_{b2} tends to increase the magnitude of the analysis increment relative to the magnitude of the error to be corrected; as in previous examples, this increases the risk of degradation. In many circumstances, overestimating σ_{b2} causes a mean degradation at the unobserved location.

These examples show that overestimating the magnitude of a background-error variance or covariance tends to increase the magnitude of the analysis increment, which increases the risk that adding an observation will degrade an individual realization of an analysis. The examples also show that although Eqs. (2) and (10) produce analyses with minimum variance and maximum likelihood, they do not produce analyses that are most likely to be an improvement over the background values. One can reduce the likelihood of degrading a single realization of an analysis by underestimating the magnitude of the background-error statistics. However, this also reduces the average magnitude of the improvements. Alternatively, one could focus on improving quantities that the data-assimilation algorithm was designed to minimize, such as the mean error or the standard deviation of the error, averaged over many realizations. This, like the

* Note that when results are measured over many analysis cycles or many realizations of the analysis, one is probably more concerned with mean results (panel (b) of Fig. 14) than with individual degradations. Panel (b) of Fig. 14 is symmetric; thus, when the results are averaged over many realizations, the one-variable data-assimilation example has no preference between overestimating or underestimating the background error, at least when $\sigma_o = \sigma_b$. This lack of preference may change when the analyses are used as initial conditions for a forecast or are part of an analysis–forecast cycle.

results in section 4(c), suggests that when adding observations, it may be beneficial to evaluate and interpret the results probabilistically rather than deterministically.

6. Forecast degradations in a simple low-dimensional chaotic model

The results in section 4 show that observations can degrade analyses and forecasts in the absence of observation errors and forecast-model errors. Section 5 explores how assimilating data probabilistically can lead to analysis degradations, isolated from the effects of error growth in forecasts. In this section, we consider how error growth can lead to forecast degradations, isolated from the effects of analysis degradations. To do so, we simulate a data-assimilation system that always improves the initial conditions, then compare forecasts made using a simple low-dimensional nonlinear model.

Like those in previous sections, the simple system implemented here can be used to explore many issues. In this study, however, it is used only to demonstrate a simple point: even with a much better data-assimilation system, the distribution of improvements and degradations in Fig. 8 would become more symmetric as forecast length increases—because nonlinear error growth limits predictability in the forecasting system.

The forecast model used is the three-variable nonlinear model from Lorenz (1963). The governing equations are:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\sigma x + \sigma y, \tag{12}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -xz + rx - y, \tag{13}$$

and

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - bz. \tag{14}$$

Following Lorenz (1963), we use the parameter values $\sigma = 10$, $b = 8/3$, and $r = 28$, which result in chaotic behaviour. The equations are integrated forward in time using a fourth-order Runge-Kutta scheme and a time step of 0.05 non-dimensional units. To the extent that the chaotic behaviour of the Lorenz system is similar to that of the QG model and the real atmosphere, we can use the simple model to illustrate general properties of the more complex systems. As in the QG experiments, the forecast model is assumed to be perfect.

To run the experiments, first, the Lorenz model is run for a long period of time from arbitrary initial conditions to create a true run. Following a spin-up period, the true run is interrupted every time unit to obtain a state $\mathbf{x}_\mathrm{t}$ $(= (x_\mathrm{t}, y_\mathrm{t}, z_\mathrm{t}))$, which is analogous to a synoptic situation in the simulated QG system. $\mathbf{x}_\mathrm{t}$ is perturbed two times, once to produce a simulated background state $\mathbf{x}_\mathrm{b}$, and once to produce a simulated analysis state $\mathbf{x}_\mathrm{a}$. $\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t}$ is analogous to the error in the initial conditions prior to adding observations; $\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{t}$ is analogous to the error in the initial conditions after assimilating the added observations. Next, the Lorenz model is used to produce forecasts from the simulated background state $f(\mathbf{x}_\mathrm{b})$ and from the simulated analysis state $f(\mathbf{x}_\mathrm{a})$. Each of the forecasts is evaluated with respect to the corresponding true state $f(\mathbf{x}_\mathrm{t})$, with a root-mean-square-average norm. Finally, for each forecast, the change in error resulting from the simulated added observations is calculated according to the definition in Eq. (5). This procedure is repeated for 10^4 states $\mathbf{x}_\mathrm{t}$.

In the example shown, only the variable x is perturbed. The same perturbations are used at each time: $x_\mathrm{b} = x_\mathrm{t} + 0.1$ and $x_\mathrm{a} = x_\mathrm{t} + 0.01$. Thus, the simulated assimilation of

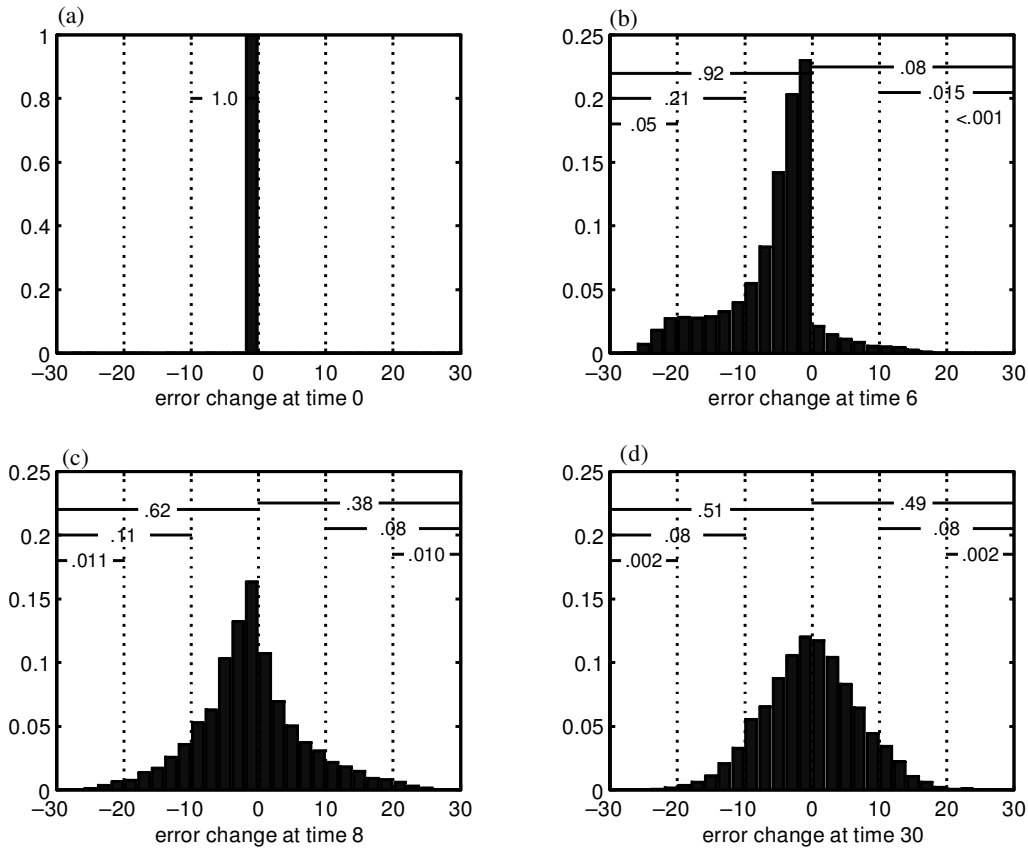

Figure 15. As in Fig. 8, but for absolute change in the error in the initial conditions (a) and in the error in forecasts of three different lengths, in the three-variable Lorenz model. The experiments are described in section 6. The error in the simulated 'initial conditions without added observations' is always the same (0.1 in x), as is the error in the simulated 'initial conditions with added observations' (0.01 in x). In other words, adding observations always reduces the initial-condition error by 90% and does not change the direction of the error.

the added observations always reduces the error in the initial conditions by 90%, without changing the direction of the error. The simulated data assimilation also improves the initial conditions in all norms. In more realistic assimilation and forecasting systems, forecast degradations can be caused by analysis degradations, or by an analysis that is improved on average but is degraded in some small rapidly growing sense. By simulating the data-assimilation system, we avoid the degradations caused by how the assimilation incorporates information at a single time, leaving only those attributable to diverging forecast trajectories.

Figure 15 depicts the histogram of Δe for forecasts of four different lengths*. One can visualize how, as the three trajectories travel around the attractor, they eventually diverge so that $f(\mathbf{x}_b)$ is at times closer to $f(\mathbf{x}_t)$ than $f(\mathbf{x}_a)$ is—this appears as a 'degradation'. How quickly a degradation appears depends on how quickly the trajectories diverge given their initial positions on the attractor. This implies that forecasts of a certain length are, in general, more likely to be degraded when the initial conditions are further apart and when errors are growing more rapidly. As one attempts to forecast

* Results are qualitatively similar for other choices of perturbations for $\mathbf{x}_b$ and $\mathbf{x}_a$; the primary difference is the rate at which degradations tend to occur.

further in advance, adding observations to the initial conditions become less and less likely to improve a deterministic forecast. At some forecast length, $f(\mathbf{x}_t)$, $f(\mathbf{x}_b)$, and $f(\mathbf{x}_a)$ are related no more closely than randomly selected states in the climatology, and degradations are as likely as improvements.

The results discussed in this section are only meant to provide a simple illustration of how forecast degradations can occur in systems that have limited predictability, even when good information is assimilated well. There are many issues that we have not addressed, including the roles of linear and nonlinear error growth in different situations and on different time-scales, and the role of model errors. As one moves to more realistic forecasting systems, these interactions become more important and the results become more complex. Many of the degradations in Fig. 8, for example, are probably caused by how the detailed structure of the analysis increment interacts with the evolution of the synoptic situation. An important next step, left for future work, is to investigate in more detail how and why observations can degrade forecasts in specific situations.

7. Increasing the likelihood and magnitude of improvement in the simulated system

As demonstrated in sections 5 and 6, adding observations can degrade analyses even if the data-assimilation system knows the error statistics perfectly, and adding observations can degrade forecasts even if the data-assimilation system always improves the initial conditions. If we cannot avoid the risk of degrading individual realizations of analyses or forecasts, then we may wish to increase the likelihood or magnitude of improvement that is expected, averaged over many realizations. In this section, we begin to explore strategies to do so by identifying circumstances in the QG and 3D-Var simulated system in which improvements are more likely and the mean change is a larger improvement.

Here, we consider primarily the role of the data-assimilation system, focusing on features of the results that are common among the cases studied and that are likely to be relevant in other data-assimilation systems and forecast models. First, we demonstrate that forecast improvement is, in general, more likely and larger when observations improve the analysis by larger amounts. Next, we show that analysis and forecast improvement is more likely and on average larger when observations are taken where ensemble spread is large, in other words where initial-condition errors are likely to be large. Then, we briefly discuss how the results depend on the spatial variation in ensemble spread at and near the observation location. Because the relationships discussed are averaged across many sets of observations, they do not necessarily hold in any specific situation. In practice, however, one can only attempt to predict expected quantities.

To help us interpret the results in this section, we use several of the concepts developed in sections 5 and 6 using low-dimensional systems. Unfortunately, the low-dimensional results cannot be extended directly to the QG and 3D-Var simulated system for several reasons. First, unlike the low-dimensional data-assimilation results presented, the 3D-Var results shown are for error-free observations. Second, 3D-Var experiments have multiple observations and many analysis locations, with correlated observation errors and correlated background errors. This, along with the error growth in the QG model, causes complex vertical, horizontal, and temporal spreading of the data, making it difficult to differentiate among causes of degradations. To account for these differences, in this section we use only those results from low-dimensional systems that are valid for error-free observations and that we believe extend to higher-dimensional

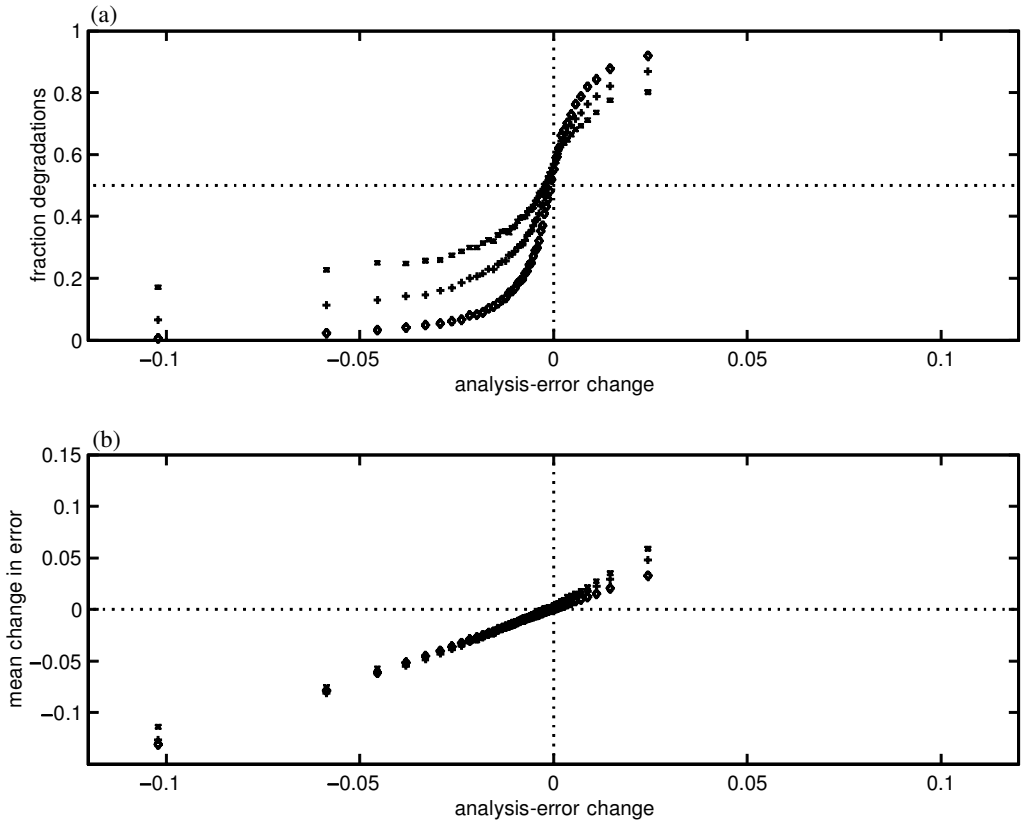

Figure 16. Forecast-error changes produced by test sets of observations that produce analysis-error changes in different ranges. Results are shown for the fraction of changes that are forecast degradations of any magnitude (a) and for the mean percent change in forecast error (b), for 12-hour forecasts ($\diamond$), 24-hour forecasts ($+$), and 48-hour forecasts ($\times$). To generate the plots, results were compiled from all of the approximately 5×10^5 tests of added observations (the same data as in Fig. 8) and divided into 50, approximately equal bins according to change in analysis error. In each bin, the fraction of forecast degradations and the mean change in forecast error were calculated, then plotted as a function of the mean change in analysis error for the bin. Each symbol plotted therefore represents a summary over approximately 10^4 test sets of observations. The few test sets of observations that produced no change in analysis error were removed. In (b), the variability of the forecast-error changes within each bin is largest for the bin with the largest analysis improvements and smallest for bins with nearly no change in analysis. The standard deviations of the forecast-error changes within the bins vary from 0.013–0.062 for 12-hour forecasts, from 0.025–0.091 for 24-hour forecasts, and from 0.040–0.123 for 48-hour forecasts.

data-assimilation systems with more than one observation. In addition, the concepts from sections 5 and 6 are used only to interpret general aspects of the high-dimensional results.

(*a*) *Observing where the data assimilation is likely to reduce analysis error*

Figure 16 illustrates the relationship between the changes in analysis error and the expected changes in forecast error when observations are added in the simulated system. Each data point represents an average over a bin containing approximately 10^4 tests of added observations; details are given in the figure caption. Figure 16 shows that added observations that improve analyses by a larger amount are more likely to improve forecasts. They also, on average, improve forecasts more. Small analysis improvements, on the other hand, tend to degrade nearly as many forecasts as they improve, and analysis

degradations tend to degrade forecasts. The link between improving the analysis and improving forecasts is only true on average, and it becomes less clear as forecast length increases.

The mechanisms for the degradations depend on the spatial structure of the errors and the error growth characteristics of the synoptic situation, neither of which are evaluated here. Thus, the results shown do not address the specific reasons why some improved analyses degrade forecasts nor why they do so on different time-scales. The results also measure only changes in the overall magnitude of the error, without accounting for changes in the direction of the error. Because small errors can grow rapidly in some directions, one could not say a priori that reducing the global-average analysis error would tend to reduce forecast errors. Figure 16 demonstrates, however, that even considering only global-average errors, large analysis improvements are much more likely than small analysis improvements (or analysis degradations) to lead to forecast improvements and to large forecast improvements.

Plotting the results with the axes switched and binned by forecast-error change (not shown) demonstrates that larger forecast improvements tend to be produced by observations that improve the analysis and that do so by larger amounts. Thus, averaged over all atmospheric situations, we are more likely to reduce the forecast error by a large amount if we can reduce the analysis error by a large amount. This suggests that, when adding observations, large analysis and forecast improvements are more likely if we understand and consider the circumstances in which the data-assimilation system is more likely (or less likely) to use additional data well.

(*b*) *Observing where initial-condition errors are likely to be large*

Recall from sections 4(a) and 4(b) that in individual synoptic situations, the analysis error tends to be reduced the most when observations are added where the ensemble spread is large. Figure 17 confirms this, averaged over all cases: where ensemble spread is larger, observations are more likely to improve analyses, and they on average improve analyses more*. Magnitude of ensemble spread at the observation location and time is also a predictor of forecast-error reduction. This suggests that observations at locations with larger ensemble spread tend to reduce not only the global-average analysis error, but also the portion of the analysis error that is important for forecasts. As in section 7(a), the relationship holds only on average, and it becomes weaker for longer-term forecasts.

Ensemble spread can be used to estimate the magnitude of initial-condition errors†. Figure 17, therefore, suggests that the 3D-Var data assimilation tends to use observational information to correct large background errors better than it corrects small background errors. This is probably caused in part by features of this 3D-Var scheme. For example, in regions with small ensemble spread, this 3D-Var scheme tends to use background-error statistics that are too large. As shown in section 5(c), overestimating background-error statistics increases the magnitude of the analysis increment, which increases the risk of degradation and decreases the mean change in error from adding observations. Incorrect background-error statistics in the 3D-Var model could, therefore,

* The unusually high likelihood of analysis improvement at locations with very small ensemble spread has not been investigated in detail. Because these locations tend to have very small initial-condition errors, however, the improvements are very small. The likelihood of degradation also increases rapidly as forecasts are made. This suggests that at locations with very small ensemble spread, assimilating added error-free observations tends to decrease the mean analysis error slightly but does not improve its spatial structure.

† Note that the ensemble, described in section 2(c), is produced by perturbing not realistic observations, but error-free observations. Thus, it cannot be constructed in the real world. For the general results discussed, however, the difference between the ensemble used here and a more realistic perturbed observation ensemble is small.

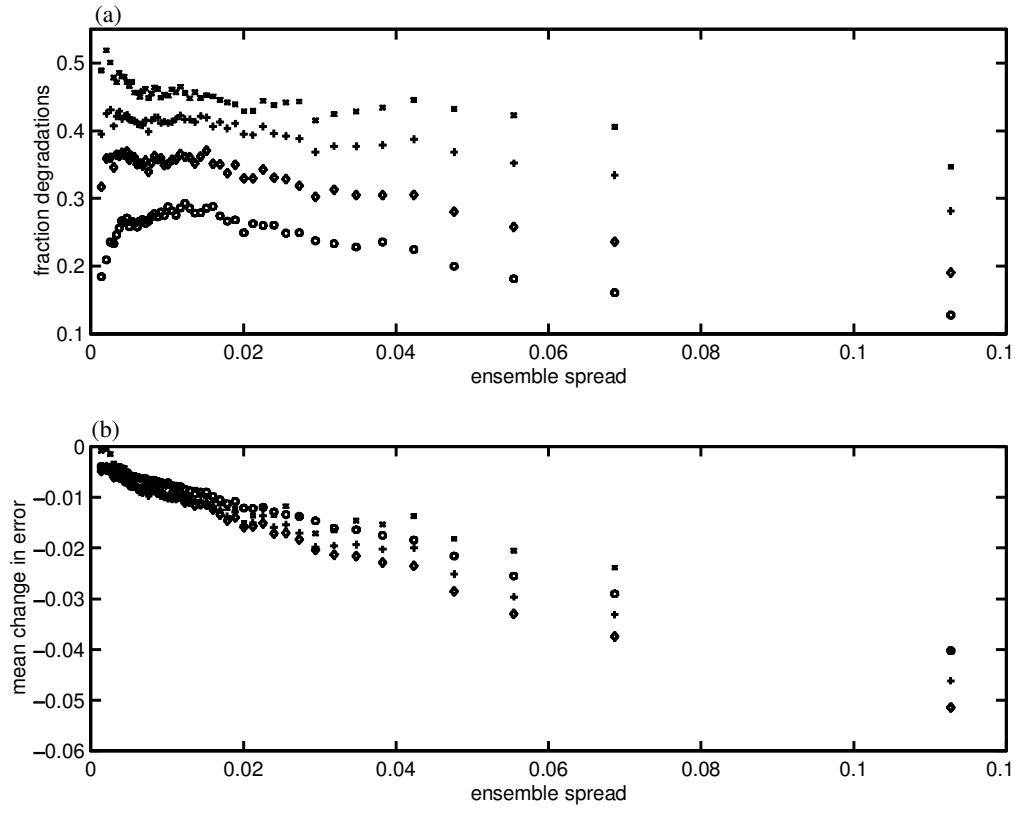

Figure 17. As in Fig. 16, for results separated into 50 approximately equal bins of ensemble spread at the test observation location and time. Results are shown for changes in errors in analyses (○), 12-hour forecasts (◇), 24-hour forecasts (+), and 48-hour forecasts (×). The ensemble contains 97 12-hour forecasts valid at the observation time; the ensemble spread is averaged over the five interior levels, with an energy norm. In (b), the variability of the error changes within each bin is smallest for the bin of smallest ensemble spread and increases with increasing ensemble spread. The standard deviations within the bins vary from 0.005–0.047 for analyses, from 0.010–0.069 for 12-hour forecasts, from 0.016–0.084 for 24-hour forecasts, and from 0.024–0.108 for 48-hour forecasts.

account for the larger risk of degradation and smaller mean-error changes on the left-hand side of Fig. 17.

In regions with large ensemble spread, this 3D-Var tends to use background-error statistics that are too small. From the results in section 5(c), we expect this to cause a small likelihood of degradation and a small mean change in error on the right-hand side of Fig. 17—instead, the mean change in error is a large improvement. This suggests that the relationships in Fig. 17 are not due solely to the time- and space-invariant magnitudes of the background-error statistics in the 3D-Var scheme. Furthermore, recall from section 5(a) that, even with perfect knowledge of the error statistics, analysis improvements are more likely and the mean change is a larger improvement when observations are added where background errors tend to be large compared with observation errors. Thus, even using a data-assimilation system with better error statistics than the 3D-Var, we are likely to have more improvements and a larger mean improvement when observations are added where ensemble spread is large.

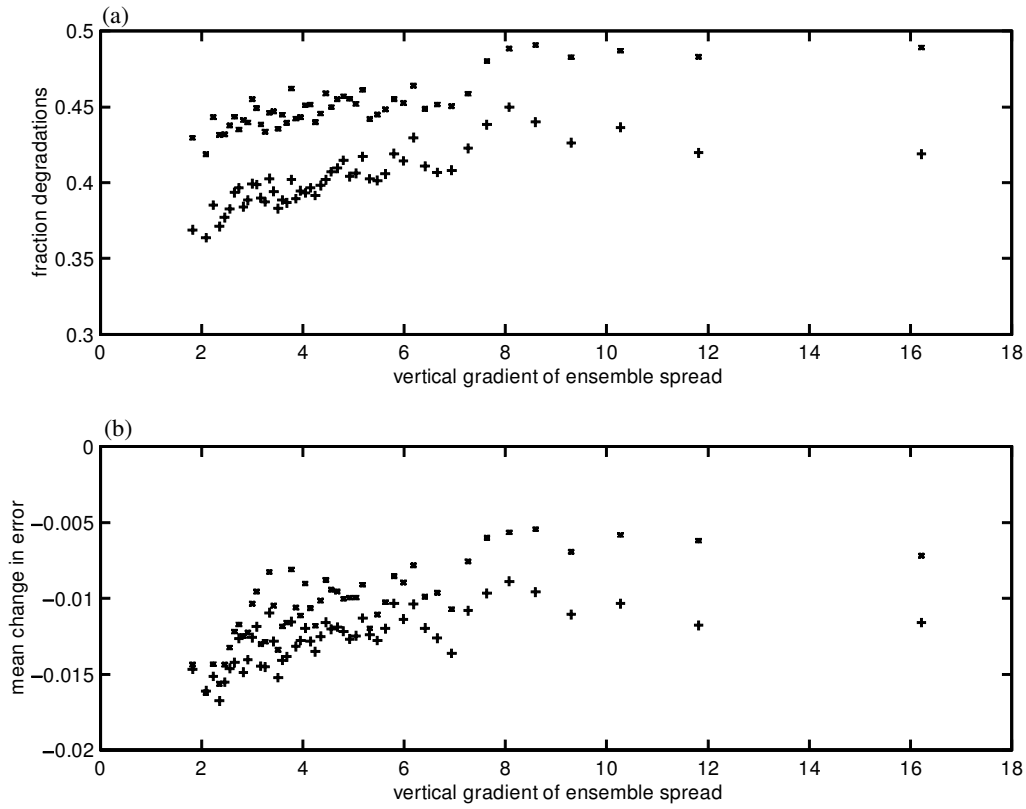

Figure 18. As in Fig. 17, for results binned by the vertical gradient of ensemble spread at observation location and time. The vertical gradient of ensemble spread is defined as the ratio between the maximum ensemble spread at any level and the minimum ensemble spread at any level. Results are shown for changes in errors in 24-hour forecasts (+) and 48-hour forecasts (×). Results for changes in errors in analyses and in 12-hour forecasts show a similar trend in the likelihood of degradation, but not in mean change in error; they are not shown to facilitate distinguishing the trend in the 24-hour and 48-hour forecasts. In (b), the variability of the error changes within each bin is large compared with the mean results; the standard deviations within the bins vary from 0.044–0.050 for 24-hour forecasts and from 0.060–0.069 for 48-hour forecasts.

(*c*) *Observing where initial-condition errors are likely to have less vertical structure*

The results in section 7(b) are for ensemble spread averaged over all levels. As Fig. 18 shows, the likelihood of improvement and the mean change in error in 1–2-day forecasts are also larger when observations are added at locations with a small vertical gradient in ensemble spread. Results for changes in errors in analyses and in 12-hour forecasts exhibit a similar trend in the likelihood of improvement, but not in mean change in error. This suggests that on average, the forecast degradations at locations with larger vertical gradient in ensemble spread are caused by somewhat different mechanisms than the degradations discussed in section 7(b), mechanisms more closely linked with how the errors grow in time.

The relationship in Fig. 18 may occur because larger vertical variation in ensemble spread tends to occur in synoptic situations that are more likely to be associated with degradations. Given the small number of cases tested, we cannot prove or disprove this hypothesis. An examination of the error changes at different levels in individual cases, however, suggests that the relationship in Fig. 18 occurs at least in part because the

data-assimilation system has difficulty correcting vertical structure in initial-condition errors. For example, at locations and times when the size of the ensemble spread is different at different levels, the analysis is often improved at levels with large spread and degraded at levels with small spread. Within 12–24 hours, this often leads to a forecast degradation at all levels. We have also noted a relationship in individual cases between horizontal variation in initial-condition errors and forecast degradations, but have not identified a parameter that shows this on average.

Our experience with the 3D-Var method supports the hypothesis that this data-assimilation system has difficulty correcting vertical structure in errors. For example, even with many observations, the 3D-Var scheme has difficulty correcting the vertical structure of initial-condition errors in terms of interior potential vorticity (Morss 1999). In general, the 3D-Var scheme has more difficulty spreading information to correct errors when the real statistics of the background errors differ greatly from the time-independent and nearly isotropic statistics assumed, such as when errors have significant spatial structure.

It is not clear how the results in Fig. 18 extend to more sophisticated data-assimilation systems. Nevertheless, the results in section 5 suggest that in any statistical data-assimilation system, with or without good knowledge of the error statistics, there are circumstances in which the likelihood of improvement and the expected magnitude of improvement are larger. Small vertical variation in ensemble spread is, at minimum, such a circumstance for this 3D-Var scheme. Thus, we propose that, when adding or adapting observations, it will be beneficial to understand the strengths and weaknesses of the data-assimilation system being used, in the context of the forecast model being used. With this knowledge, we can adjust observing strategies based on when and where the data-assimilation system is likely to perform best with respect to our goals.

8. Conclusions

The results from the simulated QG and 3D-Var system show that, even with perfect observations and a perfect forecast model, adding observations can degrade analyses and forecasts. They also provide a general idea of how likely such degradations are and how large they may be. The influence of added observations depends not only on the synoptic situation, the data-assimilation system, and the forecast model, but also on the specific locations of added observations and the specific initial-condition errors. Because of these dependencies, results in individual cases are complex and can vary significantly with the details of the experiment. This study focuses on several aspects of the results that are common among many test sets of added observations and that we believe are likely to be relevant for a variety of atmospheric data-assimilation systems and forecast models.

Deficiencies in the 3D-Var data-assimilation system contribute to many of the degradations in the simulated system. Results from experiments with two types of low-dimensional systems, however, indicate that even with a much better data-assimilation system, adding observations would still degrade some analyses and forecasts. First, low-dimensional statistical data-assimilation examples demonstrate that even with perfect knowledge of the error statistics, assimilating added observations degrades some analyses—perfect information about the statistics of errors is not equivalent to perfect information about the errors in an individual realization. Then, experiments with a low-dimensional nonlinear forecast model confirm that nonlinear error growth degrades some forecasts, even with a good forecast model and a data-assimilation system that corrects initial-condition errors well. These results suggest that the risk of degradation

is inherent in statistical data assimilation and in nonlinear prediction. They also show why many cases are needed to assess the benefits of any data-assimilation or observing strategy.

In the low-dimensional statistical data-assimilation examples, analysis degradations can have three causes: observation errors, extrapolation of data, and incorrect background-error statistics. Related to each cause, the examples identify a circumstance in which any statistical data-assimilation system is more likely to use added information to improve analyses. First, even with perfect knowledge of the error statistics, incorporating data at an observed location is more likely to improve an analysis when the observation error is likely to be small compared with the background error. Second, when extrapolating information from an observed location to an unobserved location, improvements are more likely when the mean correlation between the background errors at the two locations is larger. Third, underestimating background-error statistics produces more improvements than overestimating them. In the first two circumstances, not only are improvements more likely, but the mean change in error is also a larger improvement. In the third circumstance, improvements are more likely, but the mean change in error is small; underestimating the background-error statistics also decreases the average magnitude of analysis improvements.

The results also identify three circumstances in the simulated QG and 3D-Var system in which, averaged over all situations, adding observations is more likely to improve analyses and forecasts. First, forecasts are more likely to be improved by added observations that decrease global-average analysis errors by a larger amount. Second, analysis and forecast improvements are more likely when observations are added at locations where ensemble spread is larger, in other words where initial-condition errors are likely to be larger. Third, forecasts are more likely to be improved when observations are added where the ensemble spread has less vertical structure. In the first two circumstances, the mean change in error is also a significantly larger improvement. Further work is still needed to explore how information about the dynamical situation can be used, either alone or in conjunction with information about the likely initial-condition errors and the data-assimilation system, to identify circumstances that depend on the synoptic situation.

Together, these results provide strong evidence that if we incorporate an understanding of how the data-assimilation system and a forecast model behave into observing strategies, we can increase the likelihood of improvement and the expected improvement from adding or adapting observations. They also suggest several preliminary, simple strategies for doing so across a variety of situations. One example is to observe where the observation errors are likely to be small compared with the errors the observations are supposed to correct. The circumstances discussed in this paper are neither optimal nor exhaustive, but they are a first step towards understanding situations in which adding observations is more likely to improve analyses and forecasts and to improve them by a large amount.

Overall, the results indicate that when adding observations to improve specific deterministic forecasts, analysis and forecast degradations cannot be avoided. This suggests that assimilating added data, and using the resulting initial conditions to produce forecasts, are better interpreted as probabilistic rather than deterministic processes. If one wishes to improve a single realization of a forecast, however, the results identify several features of statistical data-assimilation systems and nonlinear forecast models that may be important to incorporate into strategies to select locations for added observations. Thus, the study not only helps us understand how adding observations might influence analysis and forecast errors in a variety of data-assimilation systems and forecast

models, but it also raises several issues to consider when developing future observing, data-assimilation, and forecasting systems.

Acknowledgements

The authors thank Chris Snyder for many useful suggestions and discussions throughout the work and the preparation of this manuscript. We also thank Andrew Lorenc and three anonymous reviewers for beneficial comments on previous versions of the manuscript, and Gerard Roe for helpful comments and discussions while revising the manuscript. This research was supported by the National Science Foundation under grant ATM-9634239. Revisions were completed while the first author was supported by a post-doctoral fellowship from the Advanced Study Program at the National Center for Atmospheric Research.

This paper is dedicated to the memory of Jorge Molina and the memory of Constantine Giannitsis.

References

Bergman, K. H. 1979 Multivariate analysis of temperatures and winds using optimum interpolation. *Mon. Weather Rev.*, **107,** 1423–1444

Bergot, T. 1999 Adaptive observations during FASTEX: A systematic survey of upstream flights. *Q. J. R. Meteorol. Soc.*, **125,** 3271–3298

Daley, R. 1991 *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, UK

Dey, C. H. and Morone, L. L. 1985 Evolution of the National Meteorological Center global data assimilation system: January 1982–December 1983. *Mon. Weather Rev.*, **113,** 304–318

Emanuel, K. A. and Langland, R. 1998 FASTEX adaptive observations workshop meeting summary, Dec. 8, 1997, Naval Research Laboratory, Monterey, Ca. *Bull. Am. Meteorol. Soc.*, **79,** 1915–1919

Errico, R. M. and Vukicevic, T. 1992 Sensitivity analysis using an adjoint of the PSU-NCAR mesoscale model. *Mon. Weather Rev.*, **120,** 1644–1660

Gelaro, R., Langland, R., Rohaly, G. D. and Rosmond, T. E. 1999 An assessment of the singular-vector approach to targeted observing using the FASTEX dataset. *Q. J. R. Meteorol. Soc.*, **125,** 3299–3328

Hamill, T. M., Snyder, C. and Morss, R. E. 2000 Error characteristics of bred, singular vector, and perturbed observation ensemble forecasts. *Mon. Weather Rev.*, **128,** 1835–1851

Hoskins, B. J. and West, N. V. 1979 Baroclinic waves and frontogenesis, part 2. Uniform potential vorticity jet flows: Cold and warm fronts. *J. Atmos. Sci.*, **36,** 1663–1680

Houtekamer, P. L. 1993 Global and local skill forecasts. *Mon. Weather Rev.*, **121,** 1834–1846

Houtekamer, P. L. and Derome, J. 1995 Methods for ensemble prediction. *Mon. Weather Rev.*, **123,** 2181–2196

Langland, R. H. 1999 Workshop on targeted observations for extratropical and tropical forecasting. *Bull. Am. Meteorol. Soc.*, **80,** 2331–2338

Langland, R. H., Gelaro, R., Rohaly, G. D. and Shapiro, M. A. 1999a Targeted observations in FASTEX: Adjoint-based targeting procedures and data impact experiments in IOPs 17 and 18. *Q. J. R. Meteorol. Soc.*, **125,** 3241–3270

Langland, R. H., Toth, Z., Gelaro R., Szunyogh, I., Shapiro, M. A., Majumdar, S., Morss, R. E., Rohaly, G. D., Velden, C., Bond, N. and Bishop, C. 1999b The North Pacific Experiment (NORPEX-98): Targeted observations for improved North American weather forecasts. *Bull. Am. Meteorol. Soc.*, **80,** 1363–1384

Lorenc, A. C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112,** 1177–1194

Lorenz, E. N. 1963 Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20,** 130–141

Lorenz, E. N. and Emanuel, K. A. 1998 Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55,** 399–414

Morss R. E. 1999 'Adaptive observations: Idealized sampling strategies for improving numerical weather prediction'. PhD thesis, Massachusetts Institute of Technology

Morss R. E., Emanuel, K. A. and Snyder, C. 2001 Idealized adaptive observation strategies for improving numerical weather prediction. *J. Atmos. Sci.*, **58,** 210–232

Parrish, D. F. and Derber, J. 1992 The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Weather Rev.*, **120,** 1747–1763

Rotunno, R. and Bao, J.-W. 1996 A case study of cyclogenesis using a model hierarchy. *Mon. Weather Rev.*, **124,** 1051–1066

Szunyogh, I., Toth, Z., Emanuel, K. A., Bishop, C., Woolen, J., Marchok, T., Morss, R. E. and Snyder, C. 1999 Ensemble-based targeting experiments during FASTEX: The effect of dropsonde data from the Lear jet. *Q. J. R. Meteorol. Soc.*, **125,** 3189–3218

Szunyogh, I., Toth, Z., Morss, R. E., Majumdar, S., Etherton, B. and Bishop, C. H. 2000 The effect of targeted dropsonde observations during the 1999 Winter Storm Reconnaissance Program. *Mon. Weather Rev.*, **128,** 3520–3537