

Maximum-Likelihood Estimation of Forecast and Observation Error Covariance Parameters. Part II: Applications

DICK P. DEE

General Sciences Corporation, Laurel, Maryland, and NASA/Goddard Space Flight Center, Greenbelt, Maryland

GREG GASPARI, CHRIS REDDER, AND LEONID RUKHOVETS

General Sciences Corporation, Laurel, Maryland

ARLINDO M. DA SILVA

NASA/Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 19 November 1997, in final form 11 August 1998)

ABSTRACT

Three different applications of maximum-likelihood estimation of error covariance parameters for atmospheric data assimilation are described. Height error standard deviations, vertical correlation coefficients, and isotropic decorrelation length scales are estimated from rawinsonde height observed-minus-forecast residuals. Sea level pressure error standard deviations and decorrelation length scales are obtained from ship reports, and wind observation error standard deviations and forecast error stream function and velocity potential decorrelation length scales are estimated from aircraft data. These applications serve to demonstrate the ability of the method to estimate covariance parameters using multivariate data from moving observers.

Estimates of the parameter uncertainty due to sampling error can be obtained as a by-product of the maximum-likelihood estimation. By bounding this source of error, it is found that many statistical parameters that are usually presumed constant in operational data assimilation systems in fact vary significantly with time. This may well reflect the use of overly simplistic covariance models that cannot adequately describe state-dependent error components such as representativeness error. The sensitivity of the parameter estimates to the treatment of bias, and to the choice of the model representing spatial correlations, is examined in detail. Several experiments emulate an online covariance parameter estimation procedure using a sliding window of data, and it is shown that such a procedure is both desirable and computationally feasible.

1. Introduction

We described the maximum-likelihood method for estimating unknown covariance parameters from time series of observed-minus-forecast residuals in Dee and da Silva (1999, hereafter referred to as Part I). In this second paper we present three different applications of the method, involving both univariate and multivariate covariance models, and using data from stationary and moving observing systems.

The primary purpose of this study is to illustrate the flexibility of the method, as well as its limitations, and to discuss many practical aspects of implementation. On the other hand, the results we report here are interesting and useful in their own right. The forecast and observation error parameter estimates obtained during the

preparation of this paper were used at NASA's Data Assimilation Office (DAO) to configure the Physical-Space Statistical Analysis System (Cohn et al. 1998), a core component of the Goddard Earth Observing System Global Data Assimilation System (GEOS DAS) (DAO 1996). In addition, we present quantitative information about the variability of some important parameters, usually specified as constants in operational data assimilation systems.

In section 2 we revisit the classic example of covariance parameter estimation for atmospheric data assimilation: the estimation of fixed-level height error standard deviations, vertical correlation coefficients, and isotropic decorrelation length scales from rawinsonde observed-minus-forecast height residuals. In section 3 we estimate the observation error standard deviation of ship-based sea level pressure reports; the observing system is not stationary in this case. We study aircraft wind observation errors in section 4; this involves a multivariate forecast error covariance model and moving observers. Conclusions are presented in section 5.

Corresponding author address: Dr. Dick P. Dee, NASA/GSFC Data Assimilation Office, Mail Code 910.3, Greenbelt, MD 20771.
E-mail: ddee@dao.gsfc.nasa.gov

All data used in this study were obtained from a February 1995 time series of observed-minus-forecast residuals produced by an early version of the GEOS DAS. This system consists of a 2° lat \times 2.5° long, 14-level analysis coupled to a 20-level, $2^\circ \times 2.5^\circ$ model for the troposphere and lower stratosphere. A description of the general circulation model, the statistical analysis system, and the quality-control procedure applied to the data may be found in Pfaendtner et al. (1995). Throughout this paper, references to numbered sections, equations, and figures in Part I will be preceded by the notation I-; for example, section I-3a refers to section 3a of Part I.

2. Rawinsonde height residuals

We first consider the estimation of rawinsonde observation and forecast error covariance parameters from observed-minus-forecast height residuals over North America for the month of February 1995. For data at a single pressure level this represents a classic example of covariance parameter estimation for atmospheric data assimilation; see for example, Gandin (1963, section 2c) and Daley (1991, section 4c). Our primary purpose here is to test the performance of the maximum-likelihood method, and to assess the accuracy of the parameter estimates in light of various uncertainties inherent in both the models and the data.

a. Height error covariance models

We use (I-6) for modeling the rawinsonde height error covariances. This model assumes that observation errors associated with different soundings are statistically independent, and that their standard deviations are identical functions of pressure only. The latter assumption may not be adequate if instruments from more than a single manufacturer are involved. Moreover, as discussed in Part I, the model does not account for representativeness error, which is state dependent and therefore spatially correlated (Daley 1993).

Forecast height error covariances are modeled by (I-8), using the *windowed power law function* ρ_w (see appendix I-A) to represent horizontal correlations. We take $r_* = 6000$ km as the distance beyond which spatial correlations are assumed to vanish. Later we will consider alternative correlation models as well. We compute cross correlations between errors at different pressure levels using the approximation (I-9).

Combining these models with the assumption that observation errors are uncorrelated with forecast errors provides the covariance of the observed-minus-forecast height residuals:

$$[\mathbf{S}]_{ij}^{mn} = \sigma_o^{(m)} \sigma_o^{(n)} \nu_o^{(mn)} \delta(r_{ij}) + \sigma_f^{(m)} \sigma_f^{(n)} \nu_f^{(mn) \frac{1}{2}} [\rho_w(r_{ij}; L^{(m)}) + \rho_w(r_{ij}; L^{(n)})]. \quad (1)$$

The notation $[\mathbf{S}]_{ij}^{(mn)}$ means the covariance of the re-

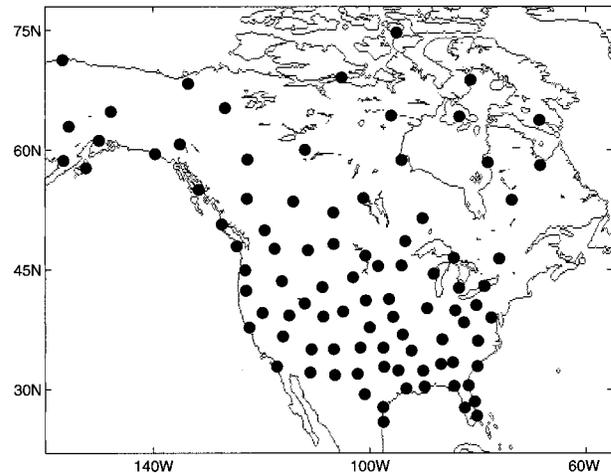


FIG. 1. Locations of North American rawinsonde stations with at least 10 simultaneous 500-hPa nighttime reports during the month of February 1995.

sidual at station i , level m with the residual at station j , level n . The quantity r_{ij} is the distance (on the surface of the earth) between stations i and j , and $\delta(r) = 1$ if $r = 0$, $\delta(r) = 0$ otherwise. The parameters $\sigma_o^{(m)}$, $\nu_o^{(mn)}$ are the observation height error standard deviations and vertical correlations, and $\sigma_f^{(m)}$, $L^{(m)}$, $\nu_f^{(mn)}$ are the forecast height error standard deviations, decorrelation length scales, and vertical correlations, respectively. Apart from these parameters the expression (1) completely specifies the residual covariances.

b. Height data

We use observed-minus-forecast height residuals based on reports from North American rawinsonde stations at standard pressure levels. Daytime rawinsonde temperature measurements are affected by sunlight, which can cause systematic errors in reported heights. Corrections may be applied to the data in order to reduce the effects of solar radiation; the method of correction generally depends on the manufacturer of the rawinsonde equipment (Mitchell et al. 1996). In some cases corrections are applied at the source; that is, prior to communicating the reports to the operational weather centers. Other than that, no corrections were applied to the data used for this study. In order to eliminate the possible contaminating effect of this aspect of quality control, we use nighttime observations only for this study.

We consider a report to be a nighttime report when the sun is below the horizon at the nominal observing time and location of the rawinsonde station. When selecting a set of stations for covariance parameter estimation we also require that all stations in the set produce a certain minimum number (usually 10) of simultaneous nighttime reports during the period in question. Figure 1 shows, for example, the locations of those North

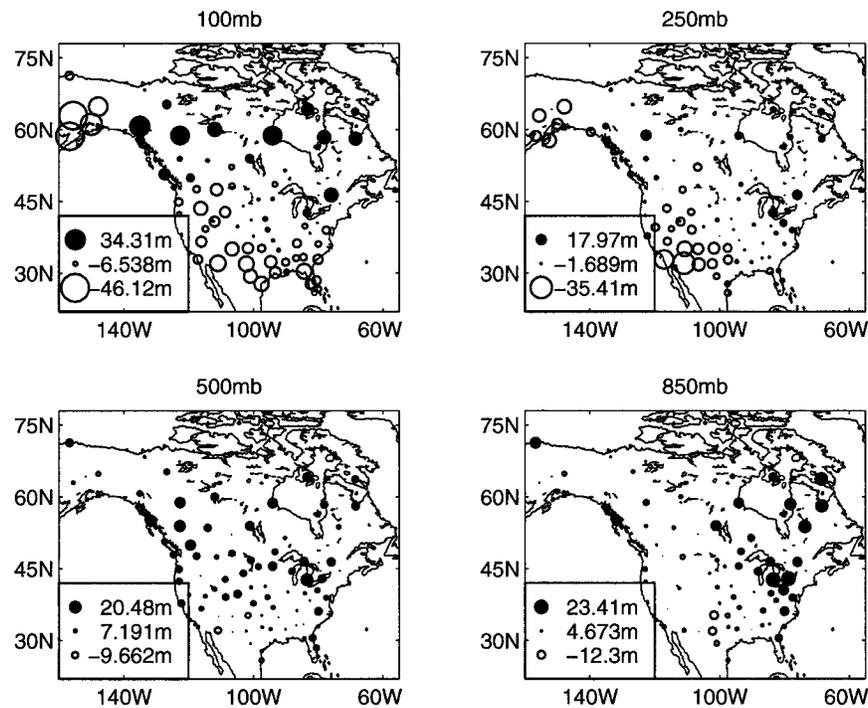


FIG. 2. Mean observed-minus-forecast nighttime height residuals for February 1995, at stations with at least 10 simultaneous nighttime reports, at 100 hPa, 250 hPa, 500 hPa, and 850 hPa. These are station averages of all nighttime residuals during the month. Closed disks correspond to positive values; open circles to negative values. The diameter of each disk or circle is proportional to the absolute value of the mean. Maximum, median, and minimum values are indicated in each panel.

American rawinsonde stations producing at least 10 simultaneous nighttime 500-hPa reports during the month of February 1995. Roughly 65% of those reports took place at 0000 UTC.

Having selected a subset of stations in this manner, we compute the mean residuals by averaging all nighttime residuals at each station; see (I-38) and (I-39). The results are shown at four different pressure levels in Fig. 2. Closed disks indicate positive mean residuals; circles correspond to negative values. The diameter of each disk or circle is proportional to the absolute value of the mean; the minimum, median, and maximum values at each level are indicated in each panel.

It is likely that the occasionally large monthly mean residuals are primarily due to systematic errors in the forecast model. For example, the means at 100 hPa clearly show a large-scale spatial pattern; it is difficult to imagine that this would be due to observational bias. The magnitudes of monthly mean height residuals computed for different periods are similar, although the detailed spatial distributions generally depend on the prevailing large-scale circulation. By introducing the assumption that the observations are unbiased, it is possible to estimate forecast bias from the time series of residuals. Dee and da Silva (1998) have developed a sequential forecast bias estimation algorithm that can be incorporated into existing statistical data assimilation

systems. The algorithm will produce multivariate forecast bias estimates that can be used to reduce climate errors in assimilated datasets.

c. Estimation of fixed-level parameters

At pressure level m , (1) reduces to

$$[\mathbf{S}]_{ij}^{(mm)} = (\sigma_o^{(m)})^2 \delta(r_{ij}) + (\sigma_f^{(m)})^2 \rho_w(r_{ij}; L^{(m)}). \quad (2)$$

Using data at a single level only, the parameters $\sigma_o^{(m)}$, $\sigma_f^{(m)}$, $L^{(m)}$ can be estimated.

The dataset contains, for example, 3447 nighttime reports from 86 North American stations at 850 hPa, 3684 reports from 95 stations at 500 hPa, 3628 reports from 95 stations at 250 hPa, and 3195 reports from 84 stations at 100 hPa. The covariance model (2) does not depend on time so that the stationary form (I-33) of the maximum-likelihood cost function may be used. We therefore first estimate the sample covariance matrix $\bar{\mathbf{S}}$ from the data at each pressure level by applying (I-36), using the threshold $K_{ij} \geq 10$. Changing the threshold value to 1 (the lowest possible) or to 40 (the maximum number of reports per station being 56) did not have a significant effect on any of the parameter estimates. The reason is that the total number of data points is not significantly reduced by removing stations that report infrequently.

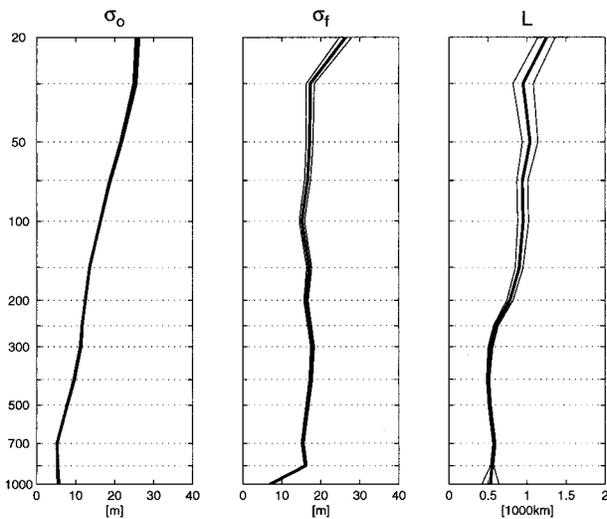


FIG. 3. Maximum-likelihood estimates, based on nighttime data only, of rawinsonde height error standard deviations (in meters), forecast height error standard deviations (in meters), and forecast height error decorrelation length scales (in thousands of kilometers), shown as a function of pressure (thick curves). Also shown are the estimated standard errors (thin curves), although these are barely visible in the leftmost panel.

Given the sample covariance $\bar{\mathbf{S}}$ at pressure level m , the log-likelihood function (I-33) can be minimized with respect to the covariance parameters $\sigma_o^{(m)}$, $\sigma_f^{(m)}$, and $L^{(m)}$. We use a quasi-Newton method with a BFGS update (Gill et al. 1981) for this purpose, allowing the scheme to approximate function gradients as needed by finite differences. Occasionally the scheme has trouble converging when the initial guess for the parameters is poor; then a simplex search method (Nelder and Mead 1965) is used instead. Optimization is considered complete when an iteration results in a relative change of less than 10^{-4} in each parameter estimate as well as in the value of the log-likelihood function. We can also compute the maximum-likelihood parameter accuracies as in (I-42), using finite differences to approximate the Hessian matrix \mathbf{A} . The entire procedure requires repeated evaluation of the log-likelihood function—typically on the order of 50–100 times in this three-parameter example, depending on the initial guess for the parameters—and takes only a few seconds on a moderately powerful desktop computer.

The maximum-likelihood parameter estimates and estimated standard errors are shown in Fig. 3. As expected, observation and forecast error standard deviations increase with height, although not monotonically. The estimated standard errors for the observation error standard deviations are roughly 2%–3% at all levels, indicating that this is the most easily identifiable parameter. The standard errors for the forecast error standard deviations increase to about 6% at higher levels. The forecast error decorrelation length scale estimates vary between 530 ± 110 km at 1000 hPa, 520 ± 20 km at 500

hPa, and 1250 ± 110 km at 20 hPa. The relatively large uncertainties in the length-scale estimates at 1000 hPa and at 20 hPa are due to the fact that there are fewer data available there.

d. Estimation of vertical correlation coefficients

Once the fixed-level covariance parameters $\sigma_o^{(m)}$, $\sigma_f^{(m)}$, $L^{(m)}$ have been estimated for each level, vertical correlation coefficients can be obtained by combining data from multiple levels. In principle it is possible to estimate all vertical correlation coefficients $\nu_o^{(mn)}$, $\nu_f^{(mn)}$ appearing in (1) simultaneously, but the optimization problem then becomes rather unwieldy. A more practical approach is to combine data from a single pair of levels and then to estimate the two vertical correlation coefficients defined for that pair. This procedure must then be repeated for each pair of levels.

Figure 4 shows the maximum-likelihood estimates of the rawinsonde and forecast height error vertical correlation coefficients obtained in this manner. Estimated standard errors were typically between 0.01 and 0.05. The vertical correlations of rawinsonde height errors are increasingly broad at high levels; this is due to the fact that the height errors at any given level involve an accumulation of thickness errors at all lower levels. The relatively large correlations between forecast errors in the midtroposphere with those in the stratosphere, as shown by the bumps in the two left-most lower panels in the figure, are not so easily explained. We suspect that this is a flow-dependent feature that is representative for the time and space domain in question. Overall, the main features of both sets of estimates agree well with earlier estimates of height error vertical correlations; compare, for instance, Figs. 3 and 10 of Lönnberg and Hollingsworth (1986).

e. Uncertainty analysis

The estimated standard errors, obtained from the asymptotic theory described in section I-4d, generally do not provide a realistic measure of the actual uncertainties in these monthly parameter estimates. The reason is that the estimates are not truly maximum-likelihood estimates, since many of the assumptions about the data that enter into the maximum-likelihood formulation are, in fact, violated. Let us introduce, for the sake of discussion, the *model hypothesis*, stating that all assumptions made in modeling the data are, in fact, satisfied. Under the model hypothesis the uncertainty in the parameter estimates is due only to the sampling error: the estimates depend on a finite number of noisy data. The effect of sampling error is different for each parameter and depends on the nature of the model. For example, the error bars on the decorrelation length scale estimates in Fig. 3 are relatively large compared to those on the estimates of the standard deviations: the data contain more useful information about the latter. Ultimately, the

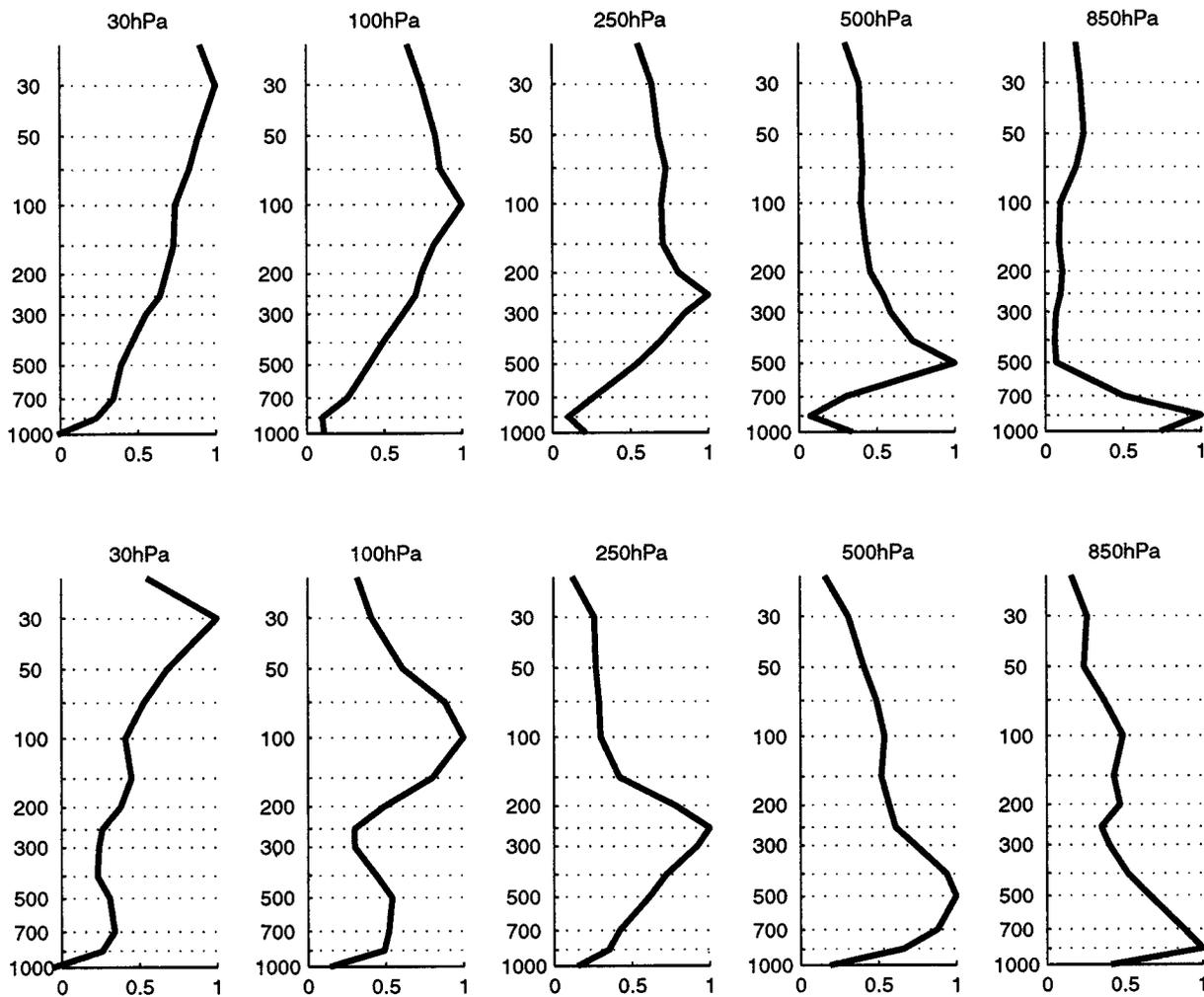


FIG. 4. Maximum-likelihood estimates at five selected pressure levels of rawinsonde height error vertical correlations (upper panels) and forecast height error vertical correlations (lower panels).

standard errors obtained from the asymptotic theory are useful because they indicate (i) whether a parameter can be actually identified from the available data and (ii) whether the parameter uncertainty due to sampling error is acceptable. The answer to both questions, for the applications presented in this study, is in the affirmative.

A better indication of the actual uncertainty in the monthly parameter estimates can be obtained by changing the data selection in various ways. The standard error estimates discussed in section 2c, for example, suggest that the sampling error would still be acceptable even if the dataset were significantly reduced. Surely it does not require many thousands of observations to estimate only three identifiable covariance parameters; this insight provided the basis for the online parameter estimation scheme proposed by Dee (1995). It also raises a number of interesting possibilities for studying the dependence of the parameter estimates on the spatial and temporal selection of data.

As an example, we emulated an online estimation

procedure by cycling through the month of February 1995 and reestimating the fixed-level covariance parameters each day, based on the most recent 10 days of data. We used nighttime rawinsonde reports only, so that the daily parameter estimates are always based on a subset (a sliding 10-day window) of the 1-month dataset described earlier. Stations with at least five simultaneous nighttime reports during the 10-day period were selected for each estimate. The number of data thus defined slowly decreases during the course of the month, due to the shortening of the night. Still, each daily estimate is based on at least 1000 observations per pressure level, which is more than sufficient for controlling sampling error. The procedure starts on 10 February, using nighttime reports from the period 1–10 February only. The resulting parameter estimates and their estimated standard errors are shown at four pressure levels as a function of time in Fig. 5.

The variability in the estimates is remarkable in several respects. The estimated observation error standard

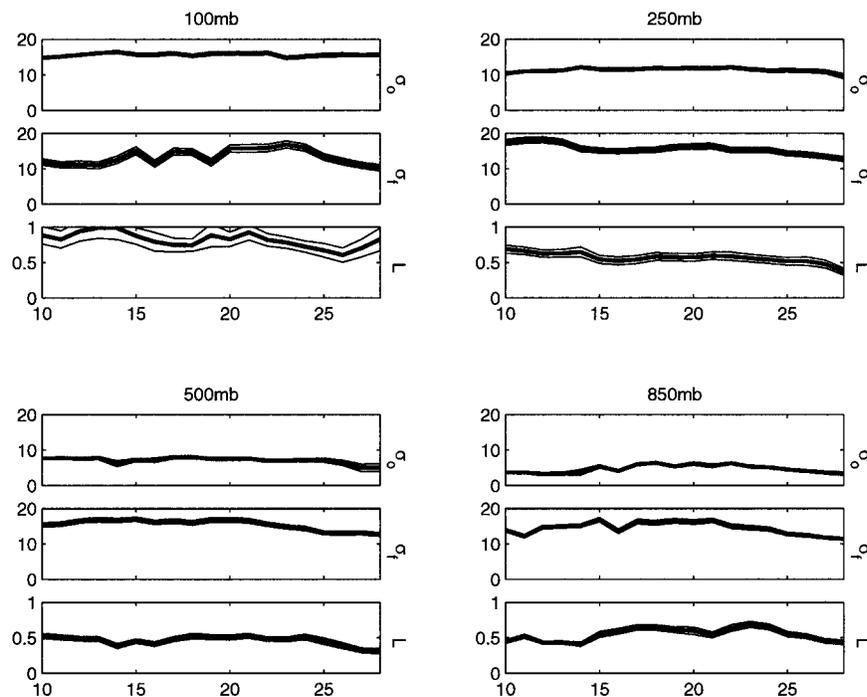


FIG. 5. Maximum-likelihood covariance parameter estimates as a function of time (in days). Shown are the estimated rawinsonde height error standard deviations (in meters), forecast height error standard deviations (in meters), and forecast height error decorrelation length scales (in thousands of kilometers) at 100 hPa, 250 hPa, 500 hPa, and 850 hPa. The estimates are produced once a day and are based on the latest 10 days of available nighttime reports. The thin curves indicate the estimated standard errors.

deviations range between 14.7 and 16.4 m at 100 hPa, between 9.6 and 12.1 m at 250 hPa, between 5.0 and 8.0 m at 500 hPa, and between 3.3 and 6.4 m at 850 hPa. Those are rather large variations for a parameter usually presumed to be a function of pressure only. The sampling error, even with a 10-day dataset, is too small to be of influence (the standard error curves for the estimates are barely visible in the figure). The variability of the estimated forecast error standard deviations is not unexpected, since forecast errors are state dependent. Note that the estimated forecast error standard deviations at 500 hPa actually exceed those at 100 hPa during most of the month. The length scale estimates at 500 hPa, for example, change from 527 km on 21 February to 313 km on 28 February.

Current operational data assimilation systems use constant (in time) parameters to describe most observation error statistics; these values are usually estimated on the basis of at least a month of data. The variability observed in the daily 10-day parameter estimates can be interpreted as an uncertainty estimate for the monthly parameter estimates: our results indicate, for example, that this uncertainty for the rawinsonde height error parameter at 500 hPa is at least 20%. This result is disturbing since this parameter, among all parameters used to describe observation error statistics, is the easiest to

estimate, and its value has a large impact on analysis accuracy.

Another way of assessing parameter uncertainty is by changing the model hypothesis and examining the effect that such a change would have on the parameter estimates. There are many dubious aspects to the model hypothesis, and a complete sensitivity analysis would not be practical. On the other hand, even a limited analysis can be valuable if it helps to identify particular components of the model formulation that can have a significant impact on the results.

Consider, for example, the specific form of the function used to represent forecast height correlations in the present application. Simultaneous estimation of observation and forecast error standard deviations from residuals is possible only by virtue of the fact that the forecast errors are spatially correlated. The precise nature of the correlations is highly uncertain and the choice of the function used to model them must have an effect on the estimated standard deviations, because the values of the likelihood function depend on this choice. To illustrate this, we repeated the parameter estimation procedure, using the same residual covariance model (2), but now with the *powerlaw* and the *compactly supported fifth-order piecewise rational* functions (see appendix I-A) representing the isotropic forecast error correla-

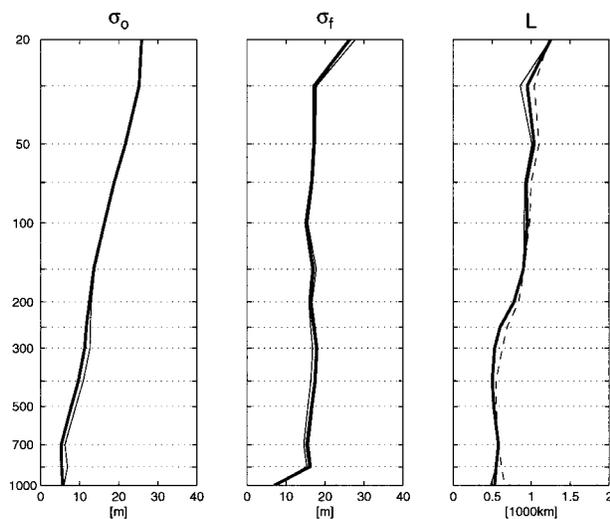


FIG. 6. The effect of the choice of forecast error correlation model on the maximum-likelihood parameter estimates. Estimates of rawinsonde height error standard deviations (in meters), forecast height error standard deviations (in meters), and decorrelation length scales (in thousands of kilometers) are shown as a function of pressure. The thick solid curves were obtained using the spline-windowed power law function (identical to Fig. 3), the thin solid curves with the power law function, and the dashed curves with the compactly supported fifth-order piecewise rational function. Parameter estimates are based on North American nighttime February 1995 data.

tions. Figure 6 shows the results. The impact on the estimates of the error standard deviations is not much larger than that of sampling error; compare Fig. 3. The decorrelation length-scale estimates do change significantly depending on the choice of correlation model, which is not surprising since this parameter describes only the behavior of the model near the origin (see Fig. I-2). We conclude that, in this case at least, and in the context of isotropic forecast error correlation models, the choice of the representing function does not significantly affect the estimates of forecast and observation error standard deviations.

An intriguing question is whether any of the isotropic correlation models considered so far actually describe the forecast errors well. The minimum values of the log-likelihood function obtained for each choice of correlation model provides some information about this: since each of the models depends on the same number (one, in this case) of parameters, the minimum of the minimum values indicates which model provides the better fit. When the parameters are tuned to the complete month of nighttime data it is found that, of the three candidate models considered, the windowed power law function consistently provides the best fit at all levels. However, this is not true when only a 10-day window of data are used. In the cycling experiment described above, there was no clear preference toward any of the three candidate models: the smallest cost function value was obtained with different models depending on the particular time and pressure level.

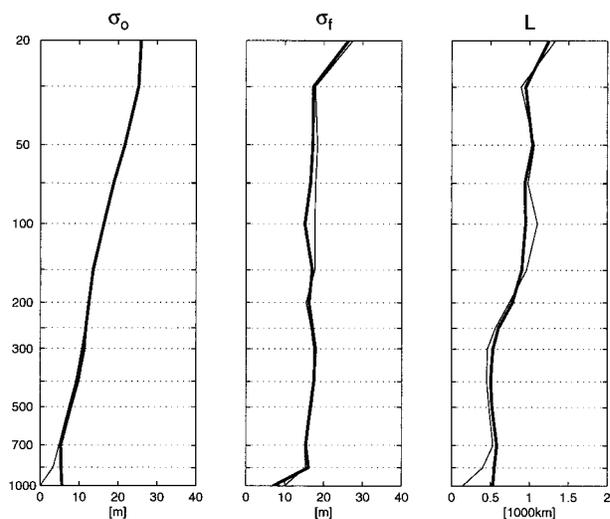


FIG. 7. The effect of the choice of estimation method on the parameter estimates. Estimates of rawinsonde height error standard deviations (in meters), forecast height error standard deviations (in meters), and decorrelation length scales (in thousands of kilometers) are shown as a function of pressure. The thick curves were obtained with the maximum-likelihood method (identical to Fig. 3); the thin curves are the Generalized Cross-Validation parameter estimates. The spline-windowed powerlaw function was used to model forecast error correlations in both cases, and the parameter estimates are based on the same set of North American nighttime February 1995 data.

It is probably the case that none of the models considered here fit the data well. Isotropic models are perhaps appropriate for describing forecast error correlations averaged over a sufficiently long (seasonal) time period, but on a shorter timescale the forecast errors are state dependent and their spatial correlations must therefore be anisotropic. Riishøjgaard (1998) has introduced a promising approach toward covariance modeling for state-dependent forecast errors. In any case, since we question the model hypothesis it would be more meaningful to study goodness-of-fit of various candidate models based on independent datasets and on a variety of parametric and nonparametric statistical tests.

f. Generalized Cross-Validation

As a final experiment with the rawinsonde data we produced parameter estimates by means of the *Generalized Cross-Validation* (GCV) method (Wahba and Wendelberger 1980), briefly described in section I-4e and summarized in appendix I-B. Figure 7 shows the GCV estimates superimposed on the maximum-likelihood estimates. The only difference between the two sets of estimates is that they are based on the minimization of two different cost functions. The estimates are not significantly different in light of the parameter uncertainties alluded to earlier, except perhaps near the surface where the GCV estimates of the observation error standard deviations are consistently smaller. The GCV estimate at 1000 hPa is zero, which is suspicious;

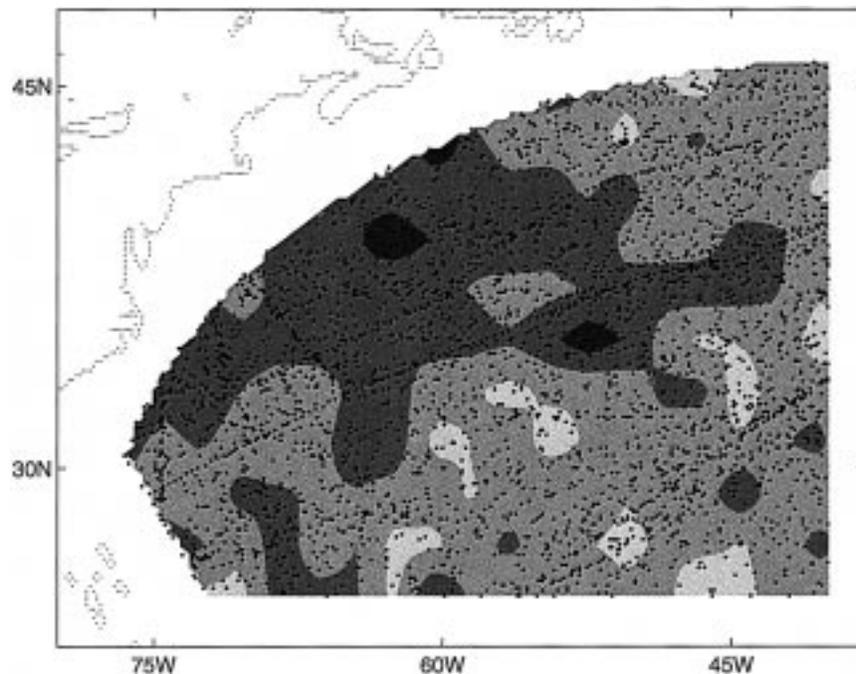


FIG. 8. Locations of February 1995 ship reports, and mean observed-minus-forecast sea level pressure residuals for that month. The mean field was computed using a successive correction method; see text for details. The four different shades of gray, from darkest to lightest, correspond to mean values in the intervals $(-3 \text{ hPa}, -2 \text{ hPa}]$, $(-2 \text{ hPa}, -1 \text{ hPa}]$, $(-1 \text{ hPa}, 0 \text{ hPa}]$, and $(0 \text{ hPa}, 1 \text{ hPa}]$.

however, it is generally not possible to determine which method is more accurate.

As discussed in section I-4e, the likelihood cost function leads to asymptotically (i.e., for large number of data) optimal parameter estimates under the model hypothesis. This property is academic if the model hypothesis is violated, as is invariably the case in practice. An important practical advantage of the maximum-likelihood method, however, is that it produces standard errors of the parameter estimates that can be interpreted as estimates of the parameter uncertainty due to sampling error. Conceptually we prefer the maximum-likelihood formulation because it is consistent with current implementations of statistical analysis systems. However, we have found the GCV method to be computationally more robust in some cases when the initial parameter estimates were very poor. In those cases the initial phase of the optimization process (the bracketing or approximate localization of the minimum) was more rapidly achieved for the GCV cost function than for the log-likelihood function. This is probably due to the fact that the GCV method first estimates the ratio of the observation and forecast error variances; see (I-45). This ratio is generally more easily identifiable from the data than each of the variances separately.

3. Sea level pressure residuals from ship reports

Next we apply the maximum-likelihood method to the estimation of sea level pressure observation and

forecast error parameters. An interesting aspect of this application is that the observing system is not stationary. Consequently, the general, time-dependent formulation (I-31) of the log-likelihood function must be used. This does not present any serious difficulties as long as the covariance between residuals at any two locations can be evaluated as a function of the parameters to be estimated. Computations are slower than in the stationary case, typically by a factor of 10 or so, depending on the size of the dataset. Still, the examples described in this section were easily calculated on a desktop computer.

a. Data and covariance models

We use ship-based sea level pressure reports obtained during February 1995 in a section of the North Atlantic situated off the east coast of the United States. Figure 8 shows the locations of each of the 3573 reports included in the dataset. Superimposed is an estimate of the monthly mean residuals, to be discussed below. The data distribution is fairly uniform, although some major shipping routes are clearly visible.

The covariance model for this case is similar to that for fixed-level rawinsonde height residuals: observation errors are assumed uncorrelated in space, and forecast errors are modeled with a simple univariate isotropic model. The covariance of the observed-minus-forecast sea level pressure residuals is therefore represented by

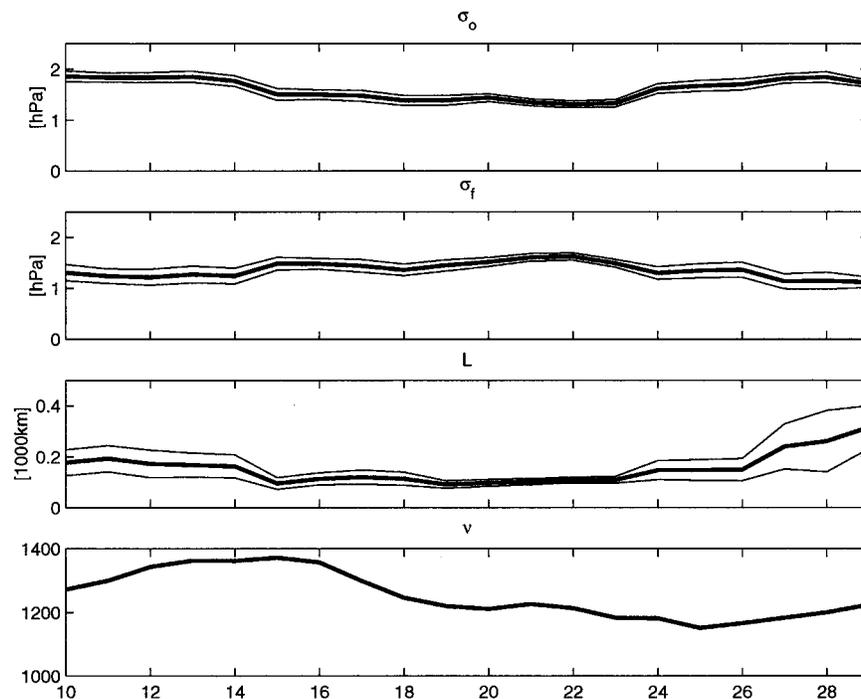


FIG. 9. Maximum-likelihood parameter estimates as a function of time (in days). The top three panels show the estimated ship sea level pressure error standard deviations (in hPa), forecast sea level pressure error standard deviations (in hPa), and forecast sea level pressure error decorrelation length scales (in thousands of kilometers). Thin curves indicate the estimated standard errors for each parameter. The estimates are produced once a day and are based on the latest 10 days of available reports. The number of reports is plotted in the bottom panel.

$$[\mathbf{S}]_{ij} = \sigma_o^2 \delta(r_{ij}) + \sigma_f^2 \rho_w(r_{ij}; L), \quad (3)$$

where σ_o and σ_f are now the observation and forecast sea level pressure error standard deviations, respectively, and L is the decorrelation length scale associated with the forecast sea level pressure errors. Given the coordinates of any two locations, this expression completely specifies the residual error covariance, except for the three parameters σ_o , σ_f , and L to be estimated from the data.

b. Bias estimation for moving observers

Bias estimation is more complicated in this case, since there are no station locations that can be used to define the bias estimates. As discussed in section I-4b, several possibilities present themselves, and we will experiment with a few of them here. The obvious approach is to construct a grid covering the data locations, and then to define and estimate the bias on this grid. (The grid may or may not coincide with the forecast model grid, but this is not relevant here). For example, the mean field shown in Fig. 8 was computed by first constructing a $2^\circ \times 2^\circ$ grid and then, for each grid location, averaging all nearest residuals. Subsequently, the estimate was smoothed by applying two iterations of a successive correction method, using a Gaussian weighting function

with a length scale of 200 km. The gridded bias estimates are then bilinearly interpolated back to the data locations for the purpose of covariance parameter estimation. Evaluation of the log-likelihood function (I-31) involves the specification of the bias at the data locations only. Hence, the bias estimates away from the observation locations do not affect the covariance parameter estimates.

The bias estimation and correction procedure just outlined is simple to implement, but involves a number of choices regarding the technical details: the definition of the grid, the method of estimation, and the interpolation scheme. Accurate bias estimates may be of interest for reasons other than covariance estimation, but our working assumption here is that *the covariance parameter estimates are not greatly sensitive to the details of the bias estimation procedure*. This assumption needs, of course, to be tested, and we will do so below. If in fact a small change in the bias estimation procedure causes a great change in the covariance parameter estimates, then the latter are not very meaningful.

Figure 9 shows, as a function of time, the maximum-likelihood estimates and estimated standard errors of the parameters σ_o , σ_f , and L based on a sliding 10-day window of data. The procedure used to obtain these estimates was described earlier in section 2. The lower

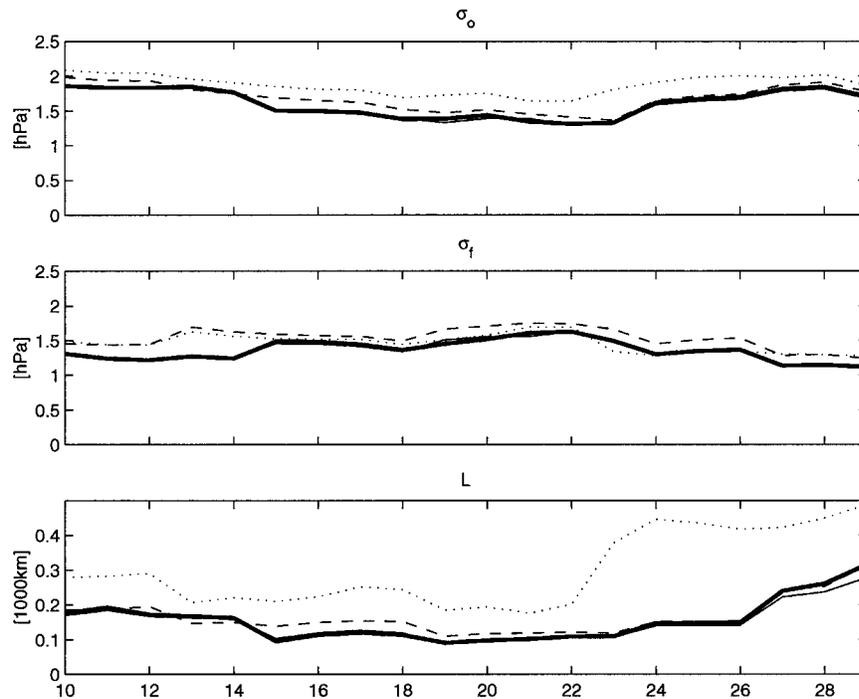


FIG. 10. The effect of the bias estimation procedure on covariance parameter estimates. The thick solid curves are identical to those in the top three panels of Fig. 9. The dotted curves were obtained by not correcting for bias at all. The thin solid curves (mostly hidden by the thick solid curves) correspond to a slight change in the bias estimation procedure (see text for details). The dashed curve was obtained by taking bias to be constant in space.

panel shows the number of data used for each estimate. Bias was estimated at each time step from the data themselves (i.e., using the same 10 days of data) on the same $2^\circ \times 2^\circ$ grid and using the same successive correction procedure described earlier in the discussion of Fig. 8.

Figure 10 shows the result of introducing various modifications to the bias estimation and correction procedure. The thick curves in this figure are identical to those in Fig. 9, while the dotted curves were obtained by not correcting for bias at all. Ignoring the bias altogether results in significantly larger decorrelation length-scale estimates and somewhat larger variance estimates. This is not surprising, since the bias in this case is mistaken for a spatially correlated random component of error. The thin solid curves are the result of using bias estimates that are less smooth: the estimation procedure involves only a single pass of the successive correction method with a more localized weighting function (using a length scale of 100 km). These curves are barely visible in the figure since they almost exactly coincide with the thick solid curves in most places. The thin dashed curves correspond to a radical simplification of the bias estimation procedure: the bias at each time was taken to be constant in space, with the constant obtained by simply averaging all the data within the 10-day period. This crude change in the procedure appears to mostly affect the estimates of the forecast error standard deviation, which turn out somewhat larger.

The differences among the parameter estimates involving some form of spatially variable bias estimation and correction are generally small; in fact, they appear to be comparable to the standard error estimates plotted in Fig. 9. This implies that, in this case at least, the details of the bias correction procedure do not significantly affect the covariance parameter estimates.

4. Wind residuals from aircraft reports

Our third application involves aircraft wind data, and the estimation of observation wind error standard deviations from these data. This is an example of parameter estimation for multivariate covariance models using moving observers.

a. Wind data and bias estimation

We used satellite-relayed wind reports obtained from various flights over a northeastern portion of the North American continent during February 1995. Reports transmitted by voice radio were excluded from the dataset. It is known that there are significant variations in the quality of reports from different airlines (Mitchell 1998, personal communication), for example, due to the different types of equipment used, but we did not take this kind of information into account.

Figure 11 shows the 1295 locations of all two-com-

ponent wind observations reported at pressure levels between 225 hPa and 275 hPa during that month. The data distribution is highly irregular and mostly concentrated about fixed flight paths.

As before, we checked the sensitivity of our results to the treatment of bias. Figure 12 shows the February 1995 mean observed-minus-forecast wind residuals computed on a $1^\circ \times 1^\circ$ grid. The mean was computed at each grid location by first averaging (over time) all nearest data and then applying two iterations of a successive correction method, using a Gaussian weighting function with a length scale of 200 km. The figure shows a coherent pattern in the residual wind directions along the major flight paths toward the northeast. The maximum residual wind speed is 8.4 m s^{-1} ; the median is 2.0 m s^{-1} .

In Fig. 13 we show the weekly means, for each of the four weeks of the month, plotted at the same scale as in the previous figure. The number of reports during each week was 353, 408, 255, and 279, respectively. Although the predominant direction of the arrows is still visible in each of the four panels, there are obvious differences as well. One might expect the covariance parameter estimates to be quite different in this case, depending on the manner in which bias is estimated and removed from the data.

b. Wind error covariance models

Aircraft wind data compose an important source of information about upper-level atmospheric flow, yet the error characteristics associated with these data are not very well known. Here we attempt to estimate only the standard deviations of the (spatially and temporally) uncorrelated component of observation error. This does not properly account for the contribution of representativeness error, which may well be highly significant in this case. Since both forecast errors and representativeness errors are likely to be state dependent and spatially correlated, it is not clear that the two can be statistically separated. This is a good example of an identifiability problem, that is, a fundamental limitation on the ability to separate observation and forecast error covariance parameters from residuals.

Aircraft wind observation error covariances are modeled simply by

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}^u & 0 \\ 0 & \mathbf{R}^v \end{bmatrix}, \quad (4)$$

with

$$[\mathbf{R}^u]_{ij} = (\sigma_o^u)^2 \delta(r_{ij}) \quad (5)$$

and

$$[\mathbf{R}^v]_{ij} = (\sigma_o^v)^2 \delta(r_{ij}). \quad (6)$$

Here σ_o^u and σ_o^v are the observation error standard deviations for the u and v components, respectively. There

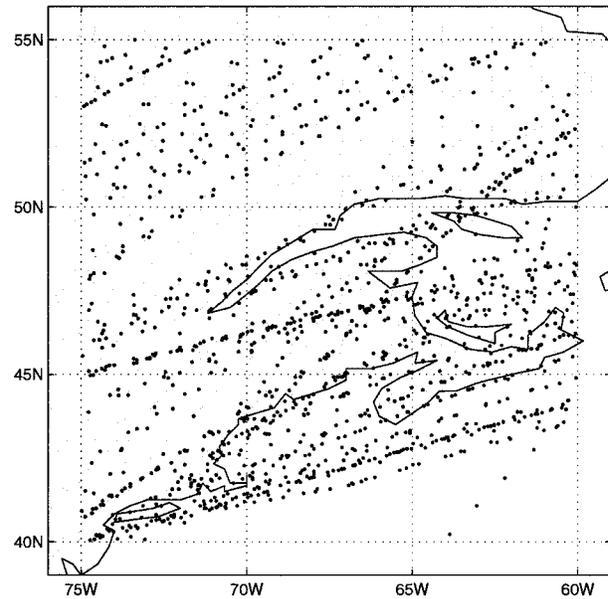


FIG. 11. Locations of February 1995 aircraft wind observations with reported pressure levels between 225 hPa and 275 hPa, within a $15^\circ \times 15^\circ$ degree portion of the North American continent.

is no obvious reason to expect that these two quantities are greatly different, but for the moment we retain the extra degree of freedom.

Wind forecast error covariances are modeled, as described in section I-2, by introducing an error streamfunction ψ and error velocity potential χ and then postulating simple univariate covariance models for each of these scalar fields. Here we assume that ψ and χ are independent with covariances

$$[\mathbf{P}^\psi]_{ij} = (\sigma^\psi)^2 \rho_w(r_{ij}; L^\psi) \quad (7)$$

and

$$[\mathbf{P}^\chi]_{ij} = (\sigma^\chi)^2 \rho_w(r_{ij}; L^\chi). \quad (8)$$

Combining (7) and (8) with (I-10) results in an anisotropic forecast wind error covariance model (see Daley 1991, his Figs. 5-2 and 5-4). In particular, the forecast wind error standard deviations σ_f^u , σ_f^v implied by this model are determined by the covariance parameters σ^ψ , L^ψ and σ^χ , L^χ .

We estimated all six parameters from the entire month of data, using the general, time-dependent formulation (I-31) of the log-likelihood function. The monthly mean shown in Fig. 12 was subtracted from the wind residuals prior to covariance tuning. The resulting parameter estimates and their estimated standard errors were

$$\sigma_o^u = (2.75 \pm 0.07) \text{ m s}^{-1}, \quad (9)$$

$$\sigma_o^v = (2.77 \pm 0.07) \text{ m s}^{-1}, \quad (10)$$

$$\sigma^\psi = (18.46 \pm 2.03) \text{ m s}^{-1}, \quad (11)$$

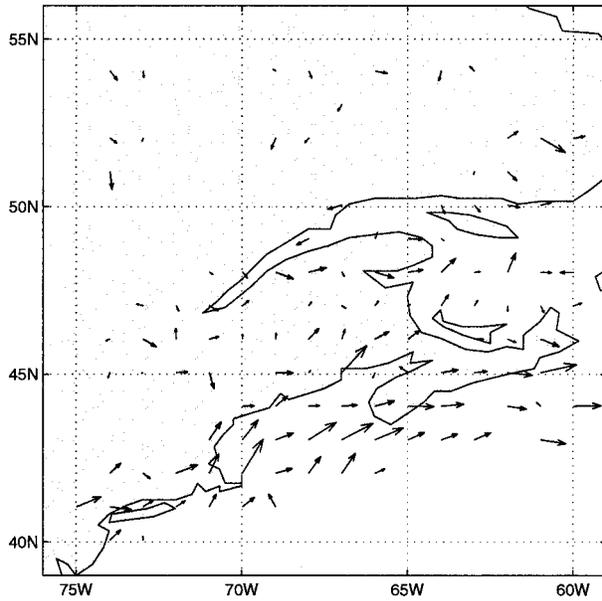


FIG. 12. Mean observed-minus-forecast wind residuals from aircraft data, computed using February 1995 reports at levels between 225–275 hPa. The maximum residual wind speed is 8.4 m s^{-1} , the median is 2.0 m s^{-1} .

$$\sigma^x = (19.77 \pm 2.59) \text{ m s}^{-1}, \quad (12)$$

$$L^\psi = (0.51 \pm 0.04) \times 10^6 \text{ m}, \quad (13)$$

and

$$L^x = (0.57 \pm 0.05) \times 10^6 \text{ m}. \quad (14)$$

The forecast wind error standard deviation implied by these estimates is $\sigma_f^u = \sigma_f^v = 3.38 \text{ m s}^{-1}$.

Note that the standard error estimates indicate that all six parameters are simultaneously identifiable from the data; the Hessian at the minimum of the log-likelihood function is well conditioned (see section I-4c).

c. Reduction of the number of parameters

The six-parameter estimation from a month of data took about an hour on a desktop workstation—the software we use for this purpose was designed for flexibility rather than efficiency. Still, it appears that the computation is unnecessarily expensive since (i) the estimates indicate that the observation error standard deviations can be represented by a single parameter and (ii) it is questionable whether four parameters are really needed to describe the forecast wind error covariance. We therefore repeated the calculation after reducing the number of parameters to three, by setting

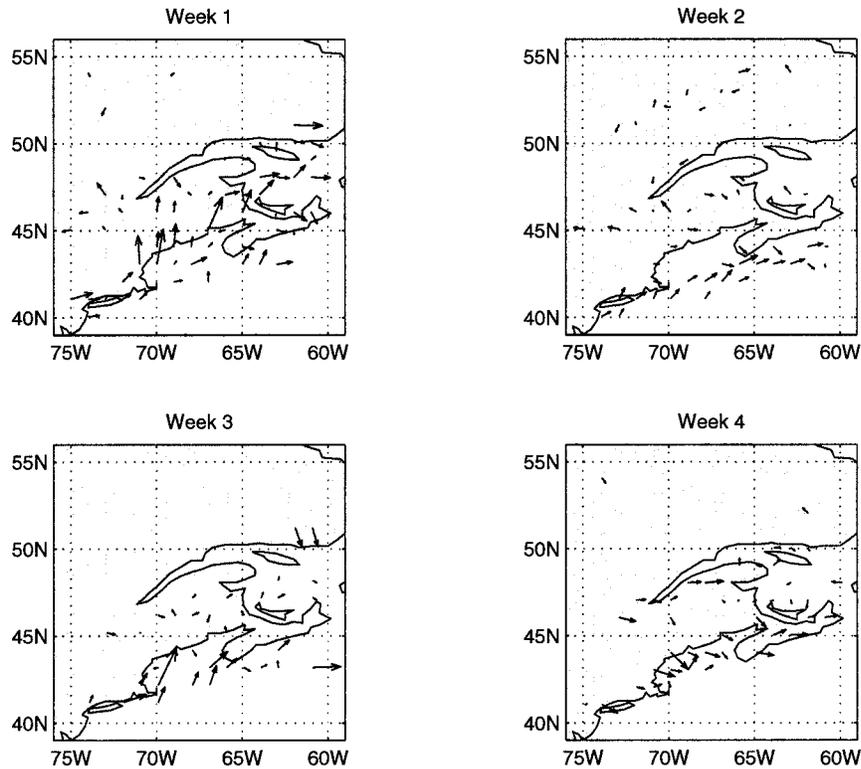


FIG. 13. As in Fig. 12, but computed for each week. The scale of the residual wind arrows in all four panels is identical to that in Fig. 12.

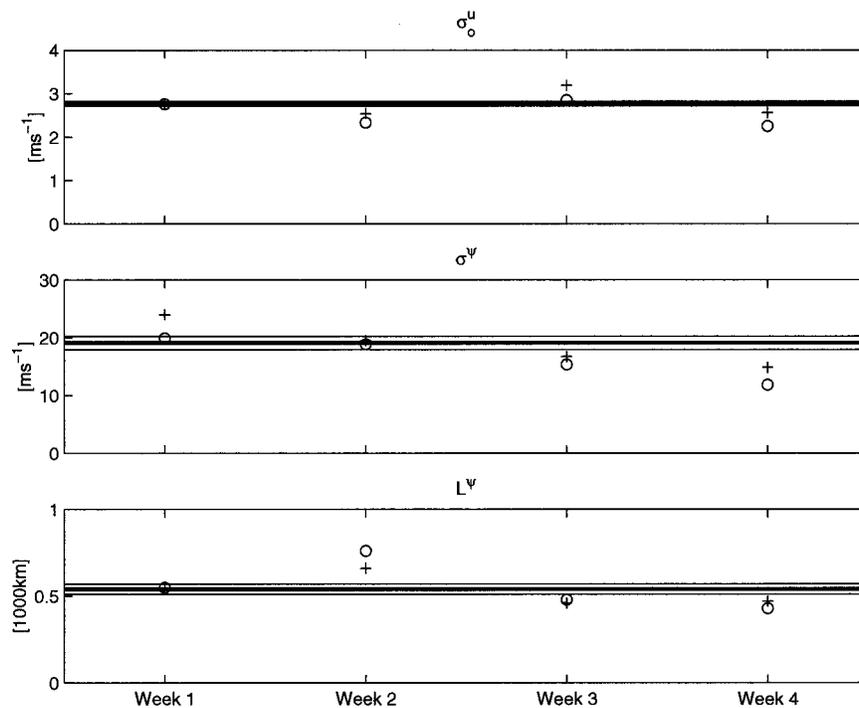


FIG. 14. Wind error covariance parameters estimated from February 1995 near 250-hPa aircraft data. Shown are the estimated aircraft observation wind error standard deviation σ_o^u (in meters per second), the forecast wind error standard deviation parameter σ^ψ (in meters per second), and the forecast wind error decorrelation length scale parameter L^ψ (in thousands of kilometers). The horizontal lines in each panel correspond to the estimates and their standard errors obtained from the entire month of data. The circles and plus signs in each panel mark weekly estimates, using two different methods for bias correction; see text for details.

$$\sigma_o^u = \sigma_o^v, \tag{15}$$

$$\sigma^\psi = \sigma^\chi, \tag{16}$$

and

$$L^\psi = L^\chi. \tag{17}$$

The resulting parameter estimates and their estimated standard errors were

$$\sigma_o^u = \sigma_o^v = (2.77 \pm 0.05) \text{ m s}^{-1}, \tag{18}$$

$$\sigma^\psi = \sigma^\chi = (19.09 \pm 1.14) \text{ m s}^{-1}, \tag{19}$$

and

$$L^\psi = L^\chi = (0.54 \pm 0.03) \times 10^6 \text{ m}. \tag{20}$$

The implied wind forecast error standard deviation is now $\sigma_f^u = \sigma_f^v = 3.37 \text{ m s}^{-1}$. It appears that in this case the wind error standard deviations can be estimated well using three parameters only.

We repeated this procedure for each week of data separately, and obtained similar results—that is, the estimated u and v observation error standard deviations are not significantly different, whether six or three parameters are used to describe the residual covariances. The estimates do vary from week to week, as shown in Fig. 14. This could be an indication of representative-

ness error, although it could also be explained by varying characteristics of the observing system; for example, different airlines reporting at different times. The horizontal lines in each panel correspond to (18)–(20); these are the estimates and their standard errors obtained from the entire month of data. The circles and plus signs in each panel mark weekly parameter estimates, using two different methods for bias correction. In the case of the circles the estimates truly depend on one week of data only: the bias was estimated from the same week of data (see Fig. 13). When, instead, the monthly mean (Fig. 12) was used as a bias estimate for each week of data, slightly different estimates were obtained (marked by the plus signs). The discrepancy between the two sets of estimates is indicative of the uncertainty due to the treatment of bias.

5. Summary and conclusions

We described three applications of the maximum-likelihood method presented in Part I for estimating forecast and observation error covariance parameters from observed-minus-forecast residuals. The first application involved nighttime data obtained from North American rawinsonde stations, used to estimate rawinsonde height error standard deviations and vertical correlations as

well as forecast height error standard deviations, decorrelation length scales, and vertical correlations. The second application, using ship-based sea level pressure reports, demonstrates the ability to estimate univariate observation error covariance parameters associated with moving observers. Parameter estimation for multivariate covariance models based on moving observers was illustrated using wind reports from aircraft.

The maximum-likelihood method produces estimates of the effect of sampling error upon parameter accuracy. By making sure that this effect is small, one can perform an uncertainty analysis of the covariance parameters by changing the data selection and/or some of the assumptions incorporated into the covariance models. For example, we studied the variability of rawinsonde height error standard deviations by reestimating them on a daily basis using only the most recent 10 days of data. We found that, within a single 1-month period, the estimated observation error standard deviations varied between 14.7 and 16.4 m at 100 hPa, between 9.6 and 12.1 m at 250 hPa, between 5.0 and 8.0 m at 500 hPa, and between 3.3 and 6.4 m at 850 hPa. A similar degree of variability was obtained in our estimates of observation error standard deviations associated with ship-based sea level pressure reports, ranging from 1.3 hPa to 1.9 hPa within the same period.

We suspect that these large variations are due to the use of overly simplistic covariance models. These models do not capture the complex, state-dependent character of the actual errors. If, for example, representativeness error is a large and significant component of the total observation error, then a great deal could be gained by explicitly accounting for it in a data assimilation system. The manner in which this could be done is an open issue in covariance modeling and deserves further study.

Other factors that can be incorporated in an uncertainty analysis of the parameter estimates include the specification of the model used to describe the spatial forecast error correlations. The results of section 2 suggest that the choice of the correlation model does not have a large impact on the parameter estimates, although this conclusion surely depends on the data distribution. However, it is probably the case that none of the models considered there, all of which are isotropic, actually fit the data well. Isotropic models are perhaps appropriate for describing forecast error correlations averaged over a sufficiently long (seasonal) time period, but on a shorter timescale the forecast errors are state dependent and their spatial correlations must therefore be anisotropic.

In some cases, monthly mean observed-minus-forecast residuals can be as large as the estimated error standard deviations themselves. Naturally this depends on the quality of the forecast model used in the assimilation system. However, particularly in case of moving observers where one cannot simply compute station means, it is necessary to be careful about the estimation and removal of bias from the data prior to the estimation

of covariance parameters. We described several experiments in sections 3 and 4 whose results indicate that the covariance parameter estimates are fairly robust with respect to the treatment of bias.

The calculations described in this paper were all performed offline with a simple desktop computer, using a stored dataset produced by an early version of GEOS DAS. Some of the experiments emulate an online covariance parameter estimation procedure, essentially as proposed by Dee (1995). Our results on the variability of the parameters support the idea that online estimation of error covariance parameters is necessary in order to take full advantage of the information contained in the observations; that is, in order to move in the direction of an optimal data assimilation method. Online estimation is computationally feasible, provided the number of data is in reasonable proportion to the number of parameters being estimated. For practical purposes this means that covariance parameters should be estimated based on local (in space and time) data only. The maximum-likelihood estimates of parameter uncertainty due to sampling error confirm that it requires a few hundreds of data items to estimate each parameter. Whether to use highly localized data spaced in time, or nearly instantaneous but spatially distributed data, is a matter of modeling strategy.

REFERENCES

- Cohn, S. E., A. M. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913–2926.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- , 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.*, **11**, 634–647.
- DAO, 1996: Algorithm theoretical basic document version 1.01. Data Assimilation Office, NASA. [Available from NASA/Goddard Space Flight Center, Greenbelt, MD 20771; or online at <http://dao.gsfc.nasa.gov/subpages/atbd.html>.]
- Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- , and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269–295.
- , and —, 1999b: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, 1822–1834.
- Gandin, L. S., 1963: *Objective Analysis of Meteorological Fields* (in Russian). Israel Program for Scientific Translation, 242 pp.
- Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical Optimization*. Academic Press, 401 pp.
- Lönnerberg, P., and A. Hollingsworth, 1986: The statistical structure of short-range forecast errors as determined from rawinsonde data. Part II: The covariance of height and wind errors. *Tellus*, **38A**, 137–161.
- Mitchell, H. L., C. Chouvard, C. Charette, R. Hogue, and S. J. Lambert, 1996: Impact of a revised analysis algorithm on an operational data assimilation system. *Mon. Wea. Rev.*, **124**, 1243–1255.
- Nelder, J. A., and R. Mead, 1965: A simplex method for function minimization. *Comput. J.*, **7**, 308–313.
- Pfaendner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz,

- J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System—Version 1. NASA Tech. Memo. 104606, Vol. 4, 44 pp. [Available from Goddard Space Flight Center, Greenbelt, MD 20771; or online at <ftp://dao.gsfc.nasa.gov/pub/techmemos/volume.4.ps.Z>.]
- Riishøjgaard, L.-P., 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A**, 42–57.
- Wahba, G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mon. Wea. Rev.*, **108**, 1122–1145.