Analysis and Quality Control of Profiler Data Using Optimum Interpolation

THOMAS NEHRKORN

Atmospheric and Environmental Research, Inc., Cambridge, Massachusetts

(Manuscript received 7 September 1998, in final form 10 August 1999)

ABSTRACT

An application of optimum interpolation (OI) is described for quality control and combination of wind profiler data with other observations of wind profiles. Data from three separate wind sensors at the National Aeronautics and Space Administration's Kennedy Space Center are used: conventional rawinsondes, precision wind sounding balloons (jimspheres), and a vertically pointing 50-MHz Doppler radar wind profiler. Collocation statistics of the three sensors are presented, along with analysis and quality control results from selected case studies. The results show the utility of the OI technique for the integration and quality control of disparate wind profile data. The proper choice of the vertical correlation length of the observation error was found to be crucial for a proper weighting of the different data sources. The performance of the OI quality control algorithm was improved if an accurate background field was used for the analysis.

1. Introduction

This paper describes an application of optimum interpolation (OI) for quality control (QC) and combination of wind profiler data with other observations of wind profiles. The method was applied to data collected at the National Aeronautics and Space Administration's (NASA's) Kennedy Space Center. The Kennedy Space Center (KSC) maintains a large array of weather observing systems to support its launch and landing operations for the space shuttle and expendable rockets. For trajectory and aerodynamic load calculations the atmospheric winds above the launch site must be determined for a period of several hours preceding the launch. Three separate observing systems are used to measure the winds: conventional rawinsondes, precision wind sounding balloons (jimspheres), and a vertically pointing 50-MHz Doppler radar wind profiler (Merceret 1997), all with different vertical and temporal resolution and error characteristics. A prototype OI algorithm described in section 2 integrates these different observations into a single analyzed wind profile, and provides an additional means of QC of the data. It is applied to selected test cases described in section 3. Results from collocation statistics of the three sensors, analysis, and QC results from the case studies are discussed in the final two sections.

2. OI methodology

We adapted an OI algorithm developed by Nehrkorn and coworkers (Nehrkorn and Hoffman 1996; Nehrkorn et al. 1997), which follows the standard OI formulation (Lorenc 1981): Analysis values at each of the grid points ($\mathbf{x}^a$) are obtained by a linear combination of the first guess or background ($\mathbf{x}^f$) and a weighted sum of the surrounding observation increments ($\mathbf{d} = \mathbf{x}^o - \mathbf{x}^f$):

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{w}^{\mathrm{T}}\mathbf{d},$$

where the weights $\mathbf{w}$ are determined from a linear equation with coefficients given in terms of the observational and background error statistics. These so-called normal equations take the form

$$\mathsf{M}\mathbf{w} = \mathbf{h}.$$

Here M is a symmetric positive definite matrix with elements given by

$$m_{ij} = \langle \pi_i \pi_j \rangle + \epsilon_i^o \langle \beta_i \beta_j \rangle \epsilon_j^o,$$

and the elements of $\mathbf{h}$ are given by

$$h_i = \langle \pi_i \pi_k \rangle.$$

In these equations, the terms $\langle \pi\pi \rangle$ are the background error correlations, and the terms $\langle \beta\beta \rangle$ are the observational error correlations, which are multiplied by ϵ^o, the ratio of observation to background error standard deviations. Correlations between background and observational errors are assumed to be zero.

For the purpose of producing a profile of analyzed wind values, we solve the analysis equations using the volume method, in which a single matrix equation is inverted for all analysis grid points within a specified

---

Corresponding author address: Dr. Thomas Nehrkorn, Atmospheric and Environmental Research, Inc., 840 Memorial Drive, Cambridge, MA 02139-3794.
E-mail: tnehrkor@aer.com

volume. The definition of an analysis volume depends on the application and may encompass all vertical levels within a given horizontal region, or it may additionally be subdivided in the vertical. In the present application there is only a single grid point in the horizontal, but a large number of vertical levels, so the analysis volumes were defined in terms of subregions in the vertical (grouping together five analysis levels at a time in the tests described below, corresponding to 1.25-km vertical subregions).[1] Correlations involving wind components are computed using the natural coordinate system of longitudinal and transverse wind components (see Daley 1991). However, since the analysis grid point (at the launch site) and the observations are essentially collocated (observing sites are within 10 km of each other, and the horizontal displacements of the sondes during ascents are typically within 40 km), the wind analysis becomes almost equivalent to performing a univariate analysis of the wind components separately. The vertical correlation functions of the background and observation errors are modeled using the functional form (see Goerss and Phoebus 1992)

$$N_{zz}(z_1, z_2) = \exp\left(-\left(\frac{|z_1 - z_2|}{dz}\right)^b\right),$$

where z_1 and z_2 are the height of the two observations, b is an adjustable constant (we used $b = 1.8$), and dz is the correlation length scale.

Quality control procedures of the OI consist of a background check and an OI buddy check, which are applied only to observations passing the QC procedures of the individual sensors (see Schumann et al. 1999 for a description of the profiler QC). In the background check, the difference between the observed value and the interpolated background value is normalized by its expected standard deviation and used in a threshold test. In the tests described below, a climatological wind profile, or an analyzed wind profile derived from an earlier balloon ascent, was used as a background field. The climatology background error standard deviations are derived from the climatological variability, and depend on location, height, and season. In the OI buddy check QC, a similar threshold test is used, except that the observed value is compared to the value obtained from an OI analysis at the observation location. In the tests described below, threshold values of 4 were used for the background QC, and 3 for the OI QC. We follow the methodology of Miller and Benjamin (1992) to determine which observations to flag as suspicious in case of large discrepancies. The OI equations are solved point by point in the QC implementation of the OI, using only a limited number of surrounding observations for each analysis point (between 4 and 12 in the tests described below). Observations to be included in the analysis are selected in a stepwise procedure designed to preferentially select observations not strongly correlated with observations already selected. In addition, only observations with errors that are uncorrelated with the observation in question (i.e., from a different sensor) are used in the OI QC analysis.

3. Case study data

The analysis and QC algorithm was tested on several case studies, representing recent launches of the space shuttle (STS83 on 4 April 1997, and STS84 on 15 May 1997), and the Titan expendable rocket (23 February 1997). The three most important wind profile data sources for those cases are the rawinsonde ascents, the jimsphere ascents, and the 50-MHz Doppler radar wind profiler. The jimsphere is a specially constructed balloon (2-m diameter) designed to oscillate with a known amplitude and frequency. The balloon ascends at a rate of 1000 ft min^{-1} (5 m s^{-1}), resulting in a 1-hr ascent to 18 km. Winds obtained from radar tracking are reported at 30-m intervals, but the measurements are considered independent at a spacing of approximately 300 m. We used a nominal observation error of 1 m s^{-1} in all our tests. Rawinsonde winds are given at a much coarser vertical spacing and are considered less accurate (we used an observation error of 2.2 m s^{-1} in our tests). Typically, the sonde ascents were spaced an hour apart, with jimsphere and rawinsonde ascents spaced a half hour apart from each other. The 50-Mhz Doppler radar wind profiler is a vertically pointing radar providing winds at range gates separated by 150 m from 2.011 to 18.6 km, with an actual vertical resolution of approximately 450 m. Median filtered profiles are available every 5 min with a root-mean-square (rms) error of approximately 1 m s^{-1} (which is the value used in our tests). A statistical comparison of the three different sensor data is presented in the next section.

The sonde and profiler data present different quality control challenges: sondes sample the atmospheric winds along a slanted path over a fairly long period and are affected by accuracy problems under extreme wind conditions when they drift far downstream; the profiler data, on the other hand, can be affected by sidelobe returns (Schumann et al. 1999 provide a detailed discussion of profiler error sources and the median filter algorithm used at KSC).

4. Collocation statistics

Wind profiles from the three different sensors were compared for the cases shown in Table 1. For purposes of this comparison, all data were interpolated vertically to the profiler levels. The rms wind vector differences between the sensors are shown for two sets of collo-

---

[1] The small vertical subregions were chosen to accommodate the high-density jimsphere observations. Alternatively, the jimsphere observations could be subsampled or combined into ‘‘superobservations.’’

TABLE 1. Observation times [in UTC, for rawinsonde (R), jimsphere (J), and profiler (P)] for the two sets of collocations and vertically averaged rms differences (in m s^{-1}).

Case	Times			rms differences		
	R	J	P	P–R	P–J	R–J
Titan	1356	1426	1506	1.38	2.67	2.38
			1426	1.58	3.05	2.39
	1456	1526	1557	1.65	1.79	1.60
			1526	1.32	1.09	1.60
	1656	1726	1758	3.27	2.79	2.31
			1727	3.20	3.23	2.31
	1756	1826	1853	3.06	2.66	2.57
			1828	2.72	2.45	2.59
	1856	1926	1953	1.79	1.69	1.72
			1928	1.49	1.66	1.72
STS83	1905	1936	1851	2.65	2.80	2.50
			1851	2.65	2.80	2.50
STS84	0738	0823	0851	2.59	2.80	2.63
			0821	2.24	2.72	2.63
All (first set)				2.83	2.70	2.52
All (second set)				2.66	2.79	2.53

cation statistics: one in which the profiler time is chosen to be roughly in the midpoint of the jimsphere ascent (first set), and another in which it is roughly coincident with the jimsphere release time (and at the midpoint of the rawinsonde ascent).[2] The vertically averaged magnitude of the differences between the jimsphere and profiler are generally consistent with results obtained by Schumann et al. (1999) for a different set of profiles. Under the assumption that the observation errors of the three sensors are uncorrelated with each other, their standard deviation (OESD) can be estimated from these difference statistics (Table 2). However, those results must be interpreted with caution for a number of reasons. The sample of cases is quite small, and there is a large variability of these statistics when computed for individual levels in the vertical, or individual cases (not shown). In addition, the sensitivity of the results to the selection of different profiler times, which results in a different ranking of the sensors for the two sets, suggests that the differences are due in some part to the space–time sampling differences between the sensors. This would imply that the rawinsonde and jimsphere differences from the true instantaneous wind profile are correlated, leading to an underestimate of their OESD and an overestimate of the profiler OESD.

5. Results

TABLE 2. Estimated observation error standard deviations (OESD, in m s^{-1}) for rawinsonde (R), jimsphere (J), and profiler (P), derived from the two sets of collocations shown in Table 1.

Case	R	J	P
First set	1.89	1.68	2.11
Second set	1.69	1.88	2.06

We computed analyzed wind values from the surface to 18 km using a vertical spacing of 250 m for this demonstration. Five analysis levels were grouped together to form 1.25-km subvolumes in the vertical. Wind observations within this volume and extending up to 20 hPa above and below the volume were considered in the OI for each volume. The OI was used to create integrated wind profiles for all times when all three sensor data were available (not shown). The influence of several adjustable parameters of the error statistics models was examined in sensitivity tests of the system. We varied the vertical correlation length scales of the background error and examined its effect on the amount of smoothing of the analysis scheme. The proper choice of the vertical correlation length of the observation error was found to be crucial, especially for the high-resolution jimsphere data, to ensure it was properly weighted relative to the other data sources. An example of the effect of varying the jimsphere observation error correlation length scale (Fig. 1) is shown in Fig. 2, which shows analyzed values for a vertical subregion for the Titan case, along with the profiler (1426 UTC) and jimsphere (1426) observations (the rawinsonde data did not extended this high for this case). The analyzed values become progressively closer to the jimsphere observations with decreasing correlation length scales, reflecting the increased weight given to the jimsphere data.

For the QC tests described below, we chose a background error correlation length scale of 800 m based on a subjective evaluation of the sensitivity tests, and observation error length scales consistent with a priori estimates (300 m for the jimsphere and rawinsonde, and 450 m for the Doppler profiler—see Fig. 1).

An example of the jimsphere and wind profiler data is shown in Fig. 3 for the case of the space shuttle launch

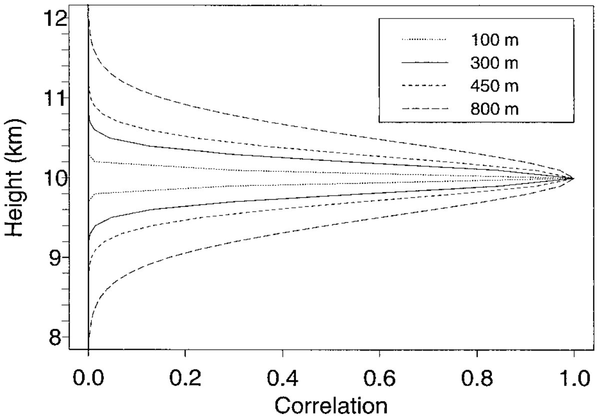

FIG. 1. Vertical correlation functions used for background (dz = 800 m) and observation errors (dz = 300 m for sonde data, dz = 450 m for profiler data).

---

[2] For STS83 the latest available profiler data was used in both sets, and the erroneous values described below were removed manually.

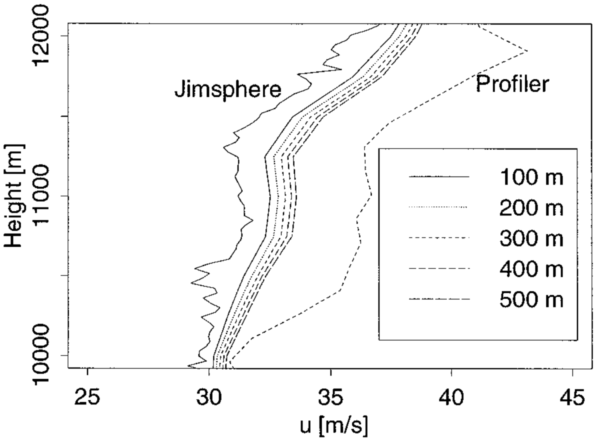

FIG. 2. Part of the zonal wind profile for the Titan case. Analyzed profiles are shown for jimsphere error correlation length scales from 100 to 500 m, alongside the jimsphere and profiler observations.

STS83 of 4 April 1997, for roughly coincident profiler (1851 UTC) and jimsphere (1750 UTC launch) times. The synoptic situation for this case is characterized by a strong upper-level trough and surface low over the Rocky Mountains, and a surface anticyclone and upper-level ridge over Florida, resulting in weak winds at the surface and northwesterly flow at upper levels over Florida. The analyzed profile in this case used a climatology background (for that time period, no observations failed the climatology background QC). The two sensors are in generally good agreement, with the notable exception of the region around 14 km, where the profiler data are clearly in error (a rawinsonde ascent at 1905 UTC corroborates the jimsphere data). A detailed view of this area is shown in Fig. 4, where it can be seen that the erroneous profiler data are all flagged by the OI QC test. However, there are several levels where the OI QC test could not unambiguously determine which observation caused the OI QC test to fail for the profiler data, and ended up flagging not only the profiler data but also some of the surrounding jimsphere data points. Nevertheless, the resulting analysis is in close agreement with the remaining jimsphere data since there are enough unflagged observations remaining to define the wind profile.

Results are quite different if a better background field is used. Using an analysis profile from one hour before as the background field causes the erroneous profiler data to be flagged by the background QC test. (Because we used an estimated background error of 1 m s^{-1} for this background, which should be considered a lower limit, compared to 11–12 m s^{-1} for climatology at 14 km, the background check is a much tighter criterion, and the analysis is much closer to the background for this case.) Even if the flagged data are permitted to enter the OI QC test (as was done in this test), however, they are again flagged by the OI QC, but this time no surrounding jimsphere are assigned OI QC flags (as shown in Fig. 5). Thus, the improved background field not only aids in the QC process directly through the background check, but also indirectly by effecting an unambiguous OI QC test result.

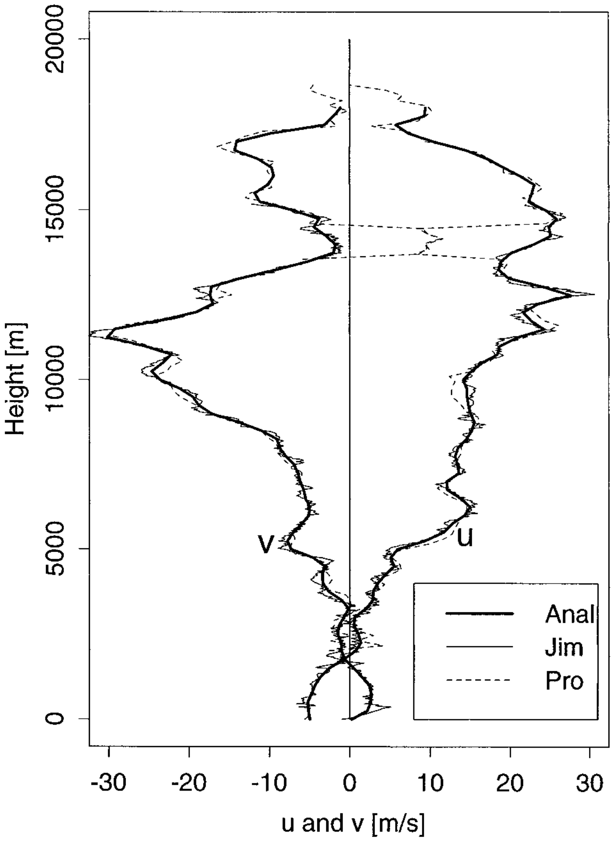

FIG. 3. Analysis (“Anal”) and observation (“Jim,” jimsphere; “Pro,” profiler) values for STS83 case.

The example shown above shows the utility of the OI technique for the integration and quality control of

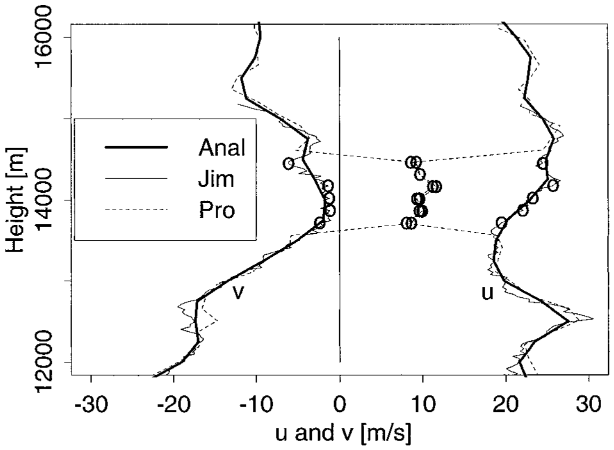

FIG. 4. Detailed view of STS83 with OI QC marks. Observations flagged by the OI QC, marked by circles, were not used to produce the analysis.

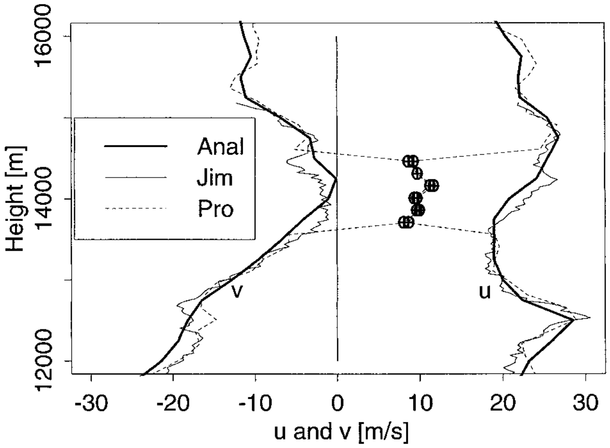

FIG. 5. The QC results with improved first guess for STS83 case. Observations flagged by the background QC (marked by plus signs) were allowed to be considered again by the OI QC and were all flagged again (marked by circles). The flagged observations were not used in the analysis.

disparate wind profile data. Other possible applications are situations where one or more wind profiler sites can be cross-checked against nearby radiosonde ascents, taking into account the different error characteristics of each sensor. The technique could be extended to take into account the horizontal drift and finite ascent time of the sondes (provided background fields with high enough horizontal and/or temporal resolutions are available). In applications where the high vertical resolution of the data is crucial, it may be desirable to increase the analysis resolution (and use background fields with shorter vertical error correlation length scales), or to only use the OI for quality control of the raw, high-resolution data.

Acknowledgments. The author benefited from helpful discussions with Dr. Ross Hoffman. Programming assistance was provided by Jeffrey Lindon. We are grateful for the case study data obtained from Dr. Merceret of the Kennedy Space Center. This research was supported by NASA Contract NAS10-97015. The original manuscript was improved through the constructive criticism of three anonymous reviewers.

REFERENCES

Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.

Goerss, J. S., and P. A. Phoebus, 1992: The navy's operational atmospheric analysis. *Wea. Forecasting,* **7,** 232–249.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

Merceret, F. J., 1997: Rapid temporal changes of midtropospheric winds. *J. Appl. Meteor.,* **36,** 1567–1575.

Miller, P. A., and S. G. Benjamin, 1992: A system for the hourly assimilation of surface observations in mountainous and flat terrain. *Mon. Wea. Rev.,* **120,** 2342–2359.

Nehrkorn, T., and R. N. Hoffman, 1996: Development of a small-scale, relocatable optimum interpolation data analysis system. *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., 19–23.

——, ——, J. Sparrow, M. Yin, S. Ryckman, and M. Leidner, 1997: Theater Analysis Procedures (TAP): Final report. PL-TR 97-2146, Phillips Laboratory, Hanscom Air Force Base, MA, 138 pp. [Available from Dr. T. Nehrkorn, Atmospheric and Environmental Research, Inc., 840 Memorial Dr., Cambridge, MA 02139-3794.]

Schumann, R. S., G. E. Taylor, F. J. Merceret, and T. L. Wilfong, 1999: Performance characteristics of the Kennedy Space Center 50-MHz Doppler wind profiler using the median filter/first-guess data reduction algorithm. *J. Atmos. Oceanic Technol.,* **16,** 532–549.