# Error Estimates for Ocean Surface Winds: Applying Desroziers Diagnostics to the Cross-Calibrated, Multiplatform Analysis of Wind Speed

ROSS N. HOFFMAN

*Atmospheric and Environmental Research, Lexington, Massachusetts*

JOSEPH V. ARDIZZONE

*NASA Goddard Space Flight Center, Greenbelt, Maryland*

S. MARK LEIDNER

*Atmospheric and Environmental Research, Lexington, Massachusetts*

DEBORAH K. SMITH

*Remote Sensing Systems, Santa Rosa, California*

ROBERT ATLAS

*NOAA/Atlantic Oceanographic and Meteorological Laboratory, Miami, Florida*

(Manuscript received 15 January 2013, in final form 3 July 2013)

ABSTRACT

The Desroziers diagnostics (DD) are applied to the cross-calibrated, multiplatform (CCMP) ocean surface wind datasets to estimate wind speed errors of the ECMWF background, the microwave satellite observations, and the resulting CCMP analysis. The DD confirm that the ECMWF operational surface wind speed error standard deviations vary with latitude in the range 0.8–1.3 m s$^{-1}$ and that the cross-calibrated Remote Sensing Systems (RSS) wind speed retrievals' standard deviations are in the range 0.5–0.7 m s$^{-1}$. Further, the estimated CCMP analysis wind speed standard deviations are in the range 0.2–0.3 m s$^{-1}$. The results suggest the need to revise the parameterization of the errors of the first guess at appropriate time (FGAT) procedure. Errors for wind speeds $<16$ m s$^{-1}$ are homogeneous; however, for the relatively rare but critical higher wind speed situations, errors are much larger.

## 1. Introduction

The cross-calibrated, multiplatform (CCMP) ocean surface wind project (Atlas et al. 2011) generates high-quality, high-resolution vector winds over the world's oceans beginning with the 1987 launch of the Special Sensor Microwave Imager (SSM/I) on *DMSP-F08*, using Remote Sensing Systems (RSS) microwave satellite wind retrievals, as well as in situ observations from ships and buoys. The CCMP data are available at the Physical Oceanography Distributed Active Archive Center (PO.DAAC; available online at http://podaac.jpl.nasa.gov/Cross-Calibrated_Multi-Platform_OceanSurfaceWindVectorAnalyses) and have been used in over 100 studies in the peer-reviewed literature in topics ranging from ocean biology (Gierach et al. 2012) to the Atlantic meridional overturning circulation (McCarthy et al. 2012), to cite two recent examples. The variational analysis method (VAM; Hoffman et al. 2003) is at the center of the CCMP project's analysis procedures for combining observations of the wind. The VAM is essentially a two-dimensional variational data assimilation. For CCMP the VAM background 10-m wind fields are from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF)

*Corresponding author address:* Dr. Ross N. Hoffman, Atmospheric and Environmental Research, 131 Hartwell Avenue, Lexington, MA 02421.
E-mail: ross.n.hoffman@aer.com

DOI: 10.1175/JTECH-D-13-00018.1

© 2013 American Meteorological Society

Re-Analysis (ERA-40) up to 1998, and from the ECMWF operational analyses thereafter. ERA-40 and ECMWF operational analyses assimilate many of the observations used in CCMP (e.g., see Uppala et al. 2005, for ERA-40). The VAM extracts additional finer spatial scales from these data (Atlas et al. 2011, Fig. SB1). The VAM was developed as a smoothing spline and so implicitly defines the background error covariance by means of several constraints with adjustable weights, and does not provide an explicit estimate of the analysis error. Eventually this work will address these two issues with the VAM—tuning the VAM in terms of the specification of observation errors and the weights used in the cost function, and assigning analysis uncertainty for the VAM products.

Here we report on our research to develop uncertainty estimates for wind speed for the VAM inputs and outputs—that is, for the background ($B$), the observations ($O$), and the analysis ($A$) wind speed—based on the Desroziers et al. (2005) diagnostics (DD). Here the DD are applied to the CCMP wind speeds for 2004. (Extensions to vector wind quantities are discussed in section 6.) The DD are applicable to any type of observation, such as satellite-observed radiances of Advanced Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (ATOVS), Atmospheric Infrared Sounder (AIRS), and Infrared Atmospheric Sounding Interferometer (IASI) (Bormann and Bauer 2010; Bormann et al. 2010, 2011). The DD have also been used within an ensemble Kalman filter (EnKF) to adaptively estimate observation errors' standard deviations and correlations (Li et al. 2009; Miyoshi et al. 2013). The DD relationships are exact only if the analysis system is optimal. In practice no analysis is truly optimal. However, information from the DD can help to iteratively refine the analysis system. The DD are only one approach to the problem of estimating errors and tuning analysis procedures. For example, the triple collocation method (e.g., Portabella and Stoffelen 2009) has also been used for this purpose. Within the smoothing splines framework (Wahba and Wendelberger 1980), one can use generalized cross validation (GCV) to estimate a small number of parameters related to the background and observation errors. However, GCV is costly, while the DD are essentially a no-cost output of analysis procedures.

## 2. The Desroziers diagnostics

Desroziers et al. (2005) describe a method to estimate background, observation, and analysis error covariances all in observation space from knowledge of the increments ($O - B$), ($A - B$), and ($O - A$). The key assumption is that the analysis is optimal. The appendix gives a derivation of the DD for error variances. Lupu et al. (2012) rewrite the equations of Desroziers et al. (2005) to show that the degree of accuracy of the estimates is related to how closely the prior covariance of ($O - B$) implied by the background and observation covariances used in the analysis matches the posterior covariance calculated from the sample. Of course no analysis is optimal, since the background and observation error statistics needed by the analysis, or the weights in the case of the VAM, are estimates and not known precisely. But important results are that the Desroziers method allows tuning of the error statistics used by the analysis and that the overall process has been shown to converge (e.g., Chapnik et al. 2006). This finding of iterative improvement suggests that while any estimates of error covariances using this method will be inexact, these will be more correct than a priori error estimates and more correct than estimates from earlier iterations.

The DD between the increments and error covariances are given in matrix form in Eqs. (2)–(4) in Desroziers et al. (2005). Here we will write the DD in scalar form for the contribution ($C$) to the estimated covariances from any two observations, denoted $i$ and $j$, as

$$C_B = (A_i - B_i)(O_j - B_j), \tag{1}$$

$$C_O = (O_i - A_i)(O_j - B_j), \tag{2}$$

$$C_A = (A_i - B_i)(O_j - A_j). \tag{3}$$

Note the following:

- The estimated covariance is the sample mean of the $C_x$ ($x = A, B, O$). The sample mean $\overline{C_x}$ may be divided into two components, one where $C_x$ has been corrected for the biases of the increments ($\overline{B_x} = \overline{c'd'}$) and a remainder due to these biases ($\overline{M_x} = \overline{c}\,\overline{d}$), where for $C_A$, for example, $c = A - B$, $d = O - A$, and a prime indicates a deviation from the mean. Results below are given in terms of $\overline{C_x}$ except where noted.
- Everything is in observation space; $A$ and $B$ are the analysis and background evaluated for the observation, at the observation time and location, respectively. The covariance estimates are for the same quantities.
- Generally $O_i$ and $O_j$ may be at different locations and different times. Results presented here are for variances ($i = j$). We will drop the subscripts $i$ and $j$ when it is possible to do so while maintaining clarity.
- The expectation operator in the DD formalism must be replaced by a sample average, in principle any

sensible sample average—all *DMSP-F14* SSM/I data within a specific ¼° grid box for one year, only those with $O > 10\,\mathrm{m\,s^{-1}}$, all Quick Scatterometer (QuikSCAT) data for the North Atlantic for the month of December in 2000–04, AMSRE data in the tropics when rain is diagnosed, etc.

- For variance estimates all the $\overline{C_x}$ should be positive. However, we find for some samples $\overline{C_A}$ is negative, and the method breaks down. See section 3 for more discussion of DD inconsistencies.
- When $i = j$, knowledge of the $C_x$ and the squared increments $(O - B)^2$, $(O - A)^2$, and $(A - B)^2$ are identical. Therefore, error variances may be estimated directly from average squared increments for any sample (see the appendix).

TABLE 1. Sample trimming by iterative QC. Sample statistics for $C_x$ for the iterative $6\sigma$ QC and values used in the VC-QC procedure. The sample size is $N = 1\,222\,254$ before VC-QC and $N = 1\,179\,420$ after VC-QC.

| QC | $i$[a] | $\sigma_B$[b] | $\sigma_O$[b] | $\sigma_A$[b] | $\mu_B$[c] | $\mu_O$[c] | $\mu_A$[c] | $f_r$[d] |
|---|---|---|---|---|---|---|---|---|
| $6\sigma$ | 1 | 6.70 | 4.04 | 2.441 | 1.94 | 0.69 | 0.171 | 0.54 |
| $6\sigma$ | 2 | 3.28 | 1.63 | 1.103 | 1.63 | 0.54 | 0.121 | 1.63 |
| $6\sigma$ | 3 | 2.49 | 1.15 | 0.814 | 1.44 | 0.46 | 0.095 | 2.46 |
| $6\sigma$ | 8 | 2.02 | 0.90 | 0.660 | 1.29 | 0.40 | 0.075 | 3.49 |
| VC[e] | — | 2.00 | 0.90 | 0.660 | 1.30 | 0.40 | 0.075 | 3.50 |

[a] The $i$ is the iteration.
[b] The $\sigma_x$ denotes sample standard deviation of $C_x$ ($\mathrm{m^2\,s^{-2}}$).
[c] The $\mu_x$ denotes sample mean of $C_x$ ($\mathrm{m^2\,s^{-2}}$).
[d] The $f_r$ is the fraction of observations rejected (%).
[e] The last row contains values used in this study for the VC-QC procedure.

## 3. Context for interpretation of DD results

One should examine the DD calculated for many different samples. When the diagnostics are inconsistent, the cause of the inconsistency can be obvious or obscure. If there were no inconsistencies, then the diagnostics confirm that the data assimilation system (DAS) is internally consistent. By refining and specializing the samples, more inconsistencies are likely to be observed. These are opportunities to discover and then mitigate errors in the formulation of the DAS. Possible reasons for inconsistencies include the following:

- Significantly non-Gaussian errors.
- Unaccounted for biases between observing systems or between observing systems and the background.
- Incorrect specification of error covariances.
- Ineffective quality control (QC) procedures.
- Lack of scale separation between observation and background errors (Desroziers et al. 2005, section 7).
- Small sample size. With one year of data, samples are generally large enough so that the formal uncertainties in $\overline{C_x}$ are quite small. However, the $C_x$ may have sample distributions that are very dispersive with heavy tails.
- Applying the DD to "observations" that were not actually used by the DAS. The VAM uses QuikSCAT as $u$ and $v$ wind components, but we use the QuikSCAT wind speeds for the DD.

Error statistics for the VAM were determined for use in the CCMP project through a series of sensitivity tests that sought an rms fit to the observations of $0.5\,\mathrm{m\,s^{-1}}$ (Atlas et al. 2011). First, satellite wind measurement errors are set to small values, with $s_u = s_v = 1\,\mathrm{m\,s^{-1}}$ and $s_w = 0.7\,\mathrm{m\,s^{-1}}$, where the subscripts $u$, $v$, and $w$ are for the two wind components and wind speed, respectively. Second, estimates of the time interpolation errors are added to the observation errors as described in Atlas et al. (2011). Third, the background constraint weights were set via the sensitivity tests. There is no explicit specification of $s_u$ and $s_v$ for the background, but these are determined from single-observation experiments to be approximately $s_u = s_v = 0.75 - 1.5\,\mathrm{m\,s^{-1}}$, varying with the synoptic situation because of the nonlinearity of the dynamical constraint used in the VAM and with latitude because of the Fourier filter used near the poles in the VAM.

## 4. Data and QC

The data used in this study are the $B$, $O$, and $A$ wind speeds from the CCMP processing for 2004. Both $B$ and $A$ are evaluated at the time of the observation according to the first guess at appropriate time (FGAT) procedure (see, e.g., Stoffelen and Anderson 1997); $B$ is interpolated linearly in time between analyses, and $A - B$ is held fixed in time over the 6-h analysis window. All values are either on or interpolated to the ¼° × ¼° grid of the RSS satellite observations. For much of what follows, we show results for a single latitude, 42.375°N.

The $C_x$ are noisy. To stabilize the estimates of $\overline{C_x}$ and their uncertainties, we have quality controlled (QC'd) the observations based on the values of $C_x$. We call this VC-QC (variance contribution QC). VC-QC is meant to be a gross QC: We eliminate observations for which at least one of the $C_x$ is more than six standard deviations from the mean. This involves an implicit definition of the sample mean and standard deviation, so we iteratively applied the $6\sigma$ test for the sample at 42.375°N until the process settled down and about 3.5% of the observations were rejected (see Table 1). We then rounded the $\sigma_x$ and $\mu_x$ estimates to obtain the values in the last row of Table 1, which are the values used in the $6\sigma$ test in what

TABLE 2. Sample statistics for all of 2004 for 42.375°N, before and after VC-QC.

| QC | $y$ | $B^a$ | $O^a$ | $A^a$ | $(O-B)^a$ | $(O-A)^a$ | $(A-B)^a$ |
|---|---|---|---|---|---|---|---|
| — | $\mu_y^b$ | 7.7 | 8.2 | 8.1 | 0.47 | 0.06 | 0.41 |
| | $\sigma_y^c$ | 3.7 | 4.1 | 4.0 | 1.55 | 0.72 | 1.27 |
| VC | $\mu_y^b$ | 7.6 | 7.9 | 7.9 | 0.34 | 0.03 | 0.31 |
| | $\sigma_y^c$ | 3.5 | 3.7 | 3.7 | 1.24 | 0.57 | 1.05 |

[a] Units are $m\,s^{-1}$.
[b] The $\mu_y$ denotes the sample mean of $y$.
[c] The $\sigma_y$ denotes the sample standard deviation of $y$.

follows. VC-QC greatly reduces the uncertainties (i.e., the size of the error bars) of our estimates of $\overline{C_x}$, and slightly reduces the estimates of $\overline{C_x}$ presented here because the data that are removed by the VC-QC tend to have higher wind speeds (Table 2). VC-QC samples are used except as noted in Tables 1 and 2.

## 5. DD results

Basic sample statistics for all platforms for 2004, for a single latitude (42.375°N; Table 2), indicate that the fit between the observations and the VAM analyses is very good—unbiased and with a standard deviation of $0.7\,m\,s^{-1}$ for this latitude. Observed wind speeds are $0.5\,m\,s^{-1}$ higher than the ECMWF background. The analysis corrects this bias and makes adjustments to the background of roughly $1.5\,m\,s^{-1}$ at locations with observations. After the VC-QC, similar results are obtained, but the magnitudes of the differences are smaller by order 25%, which is likely partly because the VC-QC has predominantly removed high wind speed cases.

DD-estimated wind speed errors for various subsamples of the 2004 data at 42.375°N are shown in Table 3. The subsamples are for different instruments and for different numbers of observations ($N_{OBS}$) used by the VAM at the particular analysis time and grid cell. The subsample $\overline{B}$ varies as much as several $0.1\,m\,s^{-1}$, and it is likely that some differences in the estimated errors are due to this inhomogeneity. In particular, the QuikSCAT sample has higher wind speeds for both $B$ and $O$ and indicates that ECMWF is biased low w.r.t. QuikSCAT and that the QuikSCAT sample includes more higher wind speed situations. In terms of the estimated observation errors ($s_O$), there is only a small variation between subsamples with $s_O^{QuikSCAT} < s_O^{AMSRE} < s_O^{SSM/I}$. As $N_{OBS}$ increases, the analysis accuracy increases from $s_A = 0.38\,m\,s^{-1}$ for $N_{OBS} = 1$ to $0.12\,m\,s^{-1}$ for $N_{OBS} = 4$. Based on the percentages for different values of $N_{OBS}$, there are on average roughly 2.5 data values contributing to the analysis in each grid cell.

Figure 1 shows how the magnitude of the estimated errors varies with latitude. The estimated error standard

TABLE 3. Sample statistics by platform and by the number of observations ($N_{OBS}$). As in Table 2. The VC-QC has been applied—sample size $N = 1\,179\,420$.

| Sample | % | $B$* | $O$* | $A$* | $C_B$** | $C_O$** | $C_A$** |
|---|---|---|---|---|---|---|---|
| All | 100 | 7.6 | 7.9 | 7.9 | 1.13 | 0.62 | 0.262 |
| QuikSCAT | 24 | 7.9 | 8.3 | 8.2 | 1.08 | 0.59 | 0.148 |
| AMSRE | 16 | 7.3 | 7.6 | 7.6 | 1.09 | 0.60 | 0.260 |
| SSM/I | 60 | 7.6 | 7.9 | 7.9 | 1.15 | 0.64 | 0.296 |
| *DMSP-F13* | 21 | 7.6 | 8.0 | 7.9 | 1.16 | 0.68 | 0.335 |
| *DMSP-F14* | 19 | 7.5 | 7.8 | 7.9 | 1.15 | 0.63 | 0.295 |
| *DMSP-F15* | 20 | 7.6 | 7.9 | 7.9 | 1.15 | 0.61 | 0.253 |
| $N_{OBS} = 1$ | 23 | 7.8 | 8.1 | 8.1 | 1.22 | 0.62 | 0.384 |
| $N_{OBS} = 2$ | 33 | 7.6 | 8.0 | 8.0 | 1.14 | 0.64 | 0.255 |
| $N_{OBS} = 3$ | 32 | 7.6 | 7.9 | 7.9 | 1.08 | 0.62 | 0.190 |
| $N_{OBS} = 4$ | 12 | 7.3 | 7.6 | 7.5 | 1.01 | 0.59 | 0.121 |

* For $B$, $O$, and $A$, the sample mean ($m\,s^{-1}$) is displayed.
** For $C_x$ the square root of the sample mean ($\sqrt{\mu_y}$, $m\,s^{-1}$) is displayed.

deviations are order 0.8–1.3 $m\,s^{-1}$ for the operational ECMWF background, 0.5–0.7 $m\,s^{-1}$ for the RSS observations, and 0.2–0.3 $m\,s^{-1}$ for the CCMP analysis. Errors tend to be smaller in the tropics, but there is a bump just north of the equator corresponding to the ITCZ. At the southernmost latitudes, the sample sizes are small, and the error estimates are anomalously large. The latter is no longer the case if the bias component is removed, as described in section 2. The bias component is large for these latitudes because $\overline{O} - \overline{B}$ (and, consequently, $\overline{A} - \overline{B}) > 1\,m\,s^{-1}$, much larger than elsewhere, perhaps because of contamination by sea ice. For comparison, Portabella and Stoffelen (2009) estimated errors using triple collocations of ECMWF analyses, *European Remote Sensing Satellite-2* (*ERS-2*) scatterometer, and 7500 buoy observations. Table 2 of Portabella and Stoffelen (2009) reports values of 2.04 and $2.16\,m\,s^{-1}$ for the estimated ECMWF wind vector standard deviation at scales of 50 km in the tropics and extratropics, respectively. Since these results are for vector errors, not speed errors, they are not inconsistent with our results. In addition, our results are nominally at a resolution of 25 km, and so contain an additional representativeness error because of variability in the 25–50-km range. Note that averaging the Portabella and Stoffelen (2009) values in an rms sense and dividing by two gives $s_u \sim s_v \sim 1.48\,m\,s^{-1}$, which is similar to the values used in the VAM (section 3).

The greatest variability for $\overline{C_x}$ is seen at high wind speeds. In Fig. 2a, for observed wind speed bins $<16\,m\,s^{-1}$, $\overline{C_x}$ is nearly constant—very slowly increasing for $B$ and $O$ and practically flat for $A$. In the two higher wind speed bins, the $\overline{C_x}$ increases very rapidly, more than doubling for the small sample of winds $>20\,m\,s^{-1}$ relative to the bins with winds $<16\,m\,s^{-1}$. First, we note that samples for the

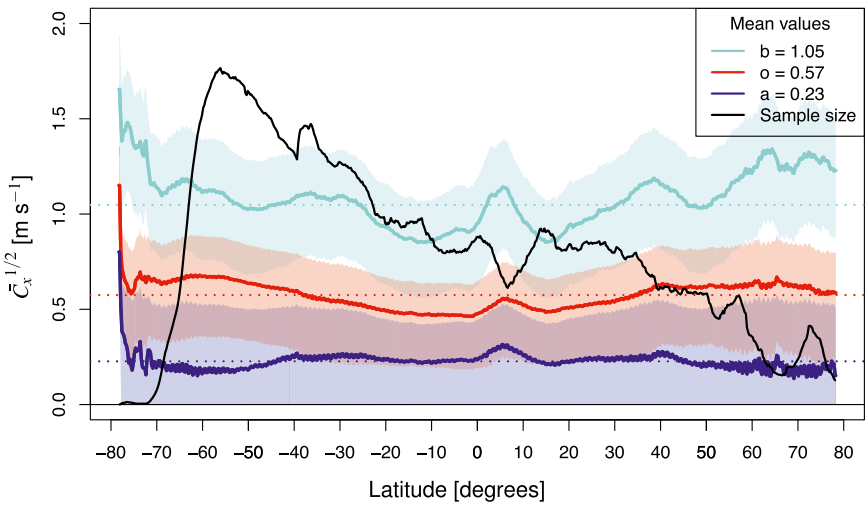


FIG. 1. Estimated error standard deviations for all of 2004, for all platforms, for each 0.25° of latitude. The colored lines show the estimated standard deviations ($\overline{C_x}^{1/2}$) for the background (aqua), observations (red), and analysis (blue). Uncertainty bands are indicated in semi-transparent hues based on ±1/3 standard deviation of the $C_x$. The overall mean values are shown in the legend and as dotted horizontal lines. The sample size divided by $2 \times 10^6$ for each 0.25° of latitude bin is plotted as a black line.

high wind speed bins are small, because this is so in nature and because the VC-QC preferentially removes high wind speeds. Fully 22% and 79% of observations in these two bins fail the VC-QC—so that only 2.32% of the QC'd sample falls in the 16–20 m s$^{-1}$ bin and only 0.24% in the >20 m s$^{-1}$ bin. Second, there are several reasons to expect larger errors at higher wind speeds. It is expected that the background errors for these cases will be large and that ECMWF will be biased low, since high wind speeds are usually associated with small-scale features and since the ECMWF data assimilation background covariances are tuned to analyze relatively large-scale features (typically >300 km). Situations with high wind speeds are more temporally and spatially variable, resulting in larger representativeness and FGAT errors that may be resulting in higher estimated errors for $O$. Also, precipitation and high ocean wave conditions often associated with high wind speeds may result in larger retrieval errors. Finally, the in situ observations used to develop and calibrate the microwave satellite retrieval algorithms are few and these algorithms might be expected to have larger errors for high wind speeds. For the analysis we expect larger analysis errors when the $B$ and/or $O$ errors are larger.

Some interesting variability for $\overline{C_x}$ is seen as the time relative to the synoptic time, $t$, varies. In Fig. 2b, for $O$, errors increase with time difference from the synoptic time. This is expected. The VAM accounts for this in the FGAT procedure, in which the background is interpolated linearly in time, and errors associated with this interpolation are considered part of the observation representativeness error. Hoffman and Leidner (2010) found 0.6–1.0 m s$^{-1}$ amplitude errors just from time interpolation over a 6-h window for wind fields from tropical cyclones in pure translation. Linear interpolation in time results in rms errors with a concave downward symmetric quadratic trend about the midpoint between synoptic times, while persistence interpolation (i.e., use the closest end point) results in an upside down V shape (Hoffman and Leidner 2010, cf. Fig. 4).

Weighted least squares fits of the binned values from Fig. 2b show no significant trends for $\overline{C_B}^{1/2}$ or $\overline{C_A}^{1/2}$, but a good linear fit for $\overline{C_O}^{1/2}$, for which the fitted values are approximately 0.53 and 0.70 m s$^{-1}$ at $t = 0$ and 180 min, respectively. Since $C_O + C_B = (O - B)^2$ as noted before, we expect that $\overline{(O - B)^2}$ will be equal to $\overline{C_O} + \overline{C_B}$. In $(O - B)^2$ we expect to have contributions from the intrinsic observation errors, the intrinsic background errors, and the FGAT interpolation errors. These are all expected to be independent so the total error variance should be the sum of the three terms, and only the FGAT error should vary with time relative to the synoptic time. Therefore, the results of the linear fits are consistent with an intrinsic observation error of 0.53 m s$^{-1}$ and a 3-h FGAT error of 0.45 m s$^{-1}$. This analysis indicates that the VAM FGAT representativeness error formulation should be revised—currently, the parabolic form is used with a 1 m s$^{-1}$ 3-h FGAT error.

Figure 3 shows the global view of the DD based on all data used by the VAM in 2004. The estimates plotted at

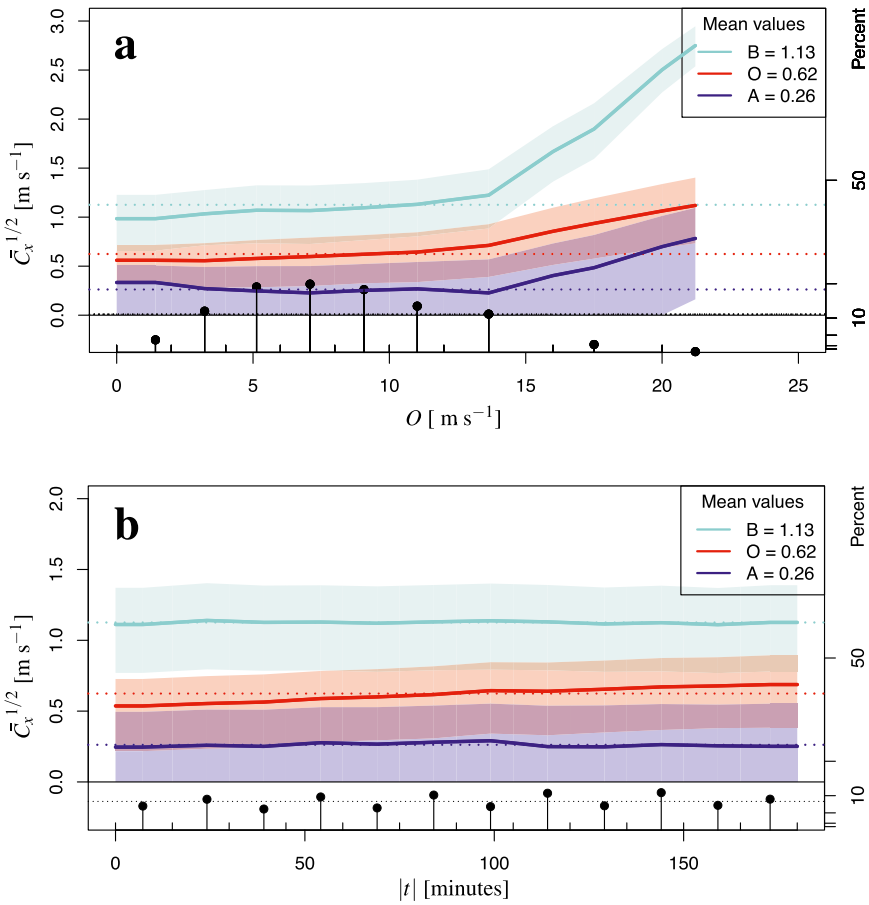


FIG. 2. Estimated error standard deviations binned by (a) observed wind speed and by (b) time relative to the synoptic time, for all of 2004, for all platforms, for 42.375°N. Plotting conventions as in Fig. 1, except that sample bin boundaries are indicated by tic marks just above the bottom axes and the percentage of the total sample in each sample bin are given by the "lollipop" symbol at the sample mean of the ordinate.

each ¼° grid box are for all observations in a centered 5 × 5 (i.e., 1.25° × 1.25°) box. This oversampling results in a smoother display. Errors in all cases are larger in the extratropics than in the tropics. Errors are largest in the western North Pacific and North Atlantic storm tracks, in the Arctic regions, and in the Southern Ocean, especially poleward of 50°. All of these are areas of enhanced storminess associated with higher wind speed and precipitation.

## 6. Concluding remarks

As a result of our investigation of applying the DD to the VAM outputs of the CCMP project for 2004, we find the following wind speed error characteristics:

- Globally, wind speed error standard deviations are estimated to vary with latitude in the range 0.8–1.3 m s$^{-1}$ for the operational ECMWF background, 0.5–0.7 m s$^{-1}$ for the RSS observations, and 0.2–0.3 m s$^{-1}$ for the CCMP analysis.
- Errors are fairly constant for observed wind speeds up to 16 m s$^{-1}$ and are much higher for higher wind speeds.
- Observational errors increase with time relative to the analysis time, and vary with platform and location. However, the variation with location may be a secondary effect of the variation with wind speed.

We plan to extend the current study in several ways. First, we will examine spatial correlations for wind speed errors in terms of distance, latitude, and longitude displacement, or in the satellite swath geometry (e.g., Bormann et al. 2011, Fig. 7). Second, we will apply the DD to two-component vector quantities—speed and direction, wind components, or pseudo-stress components. It might be advantageous here to use QuikSCAT-only analyses. Third, we will develop

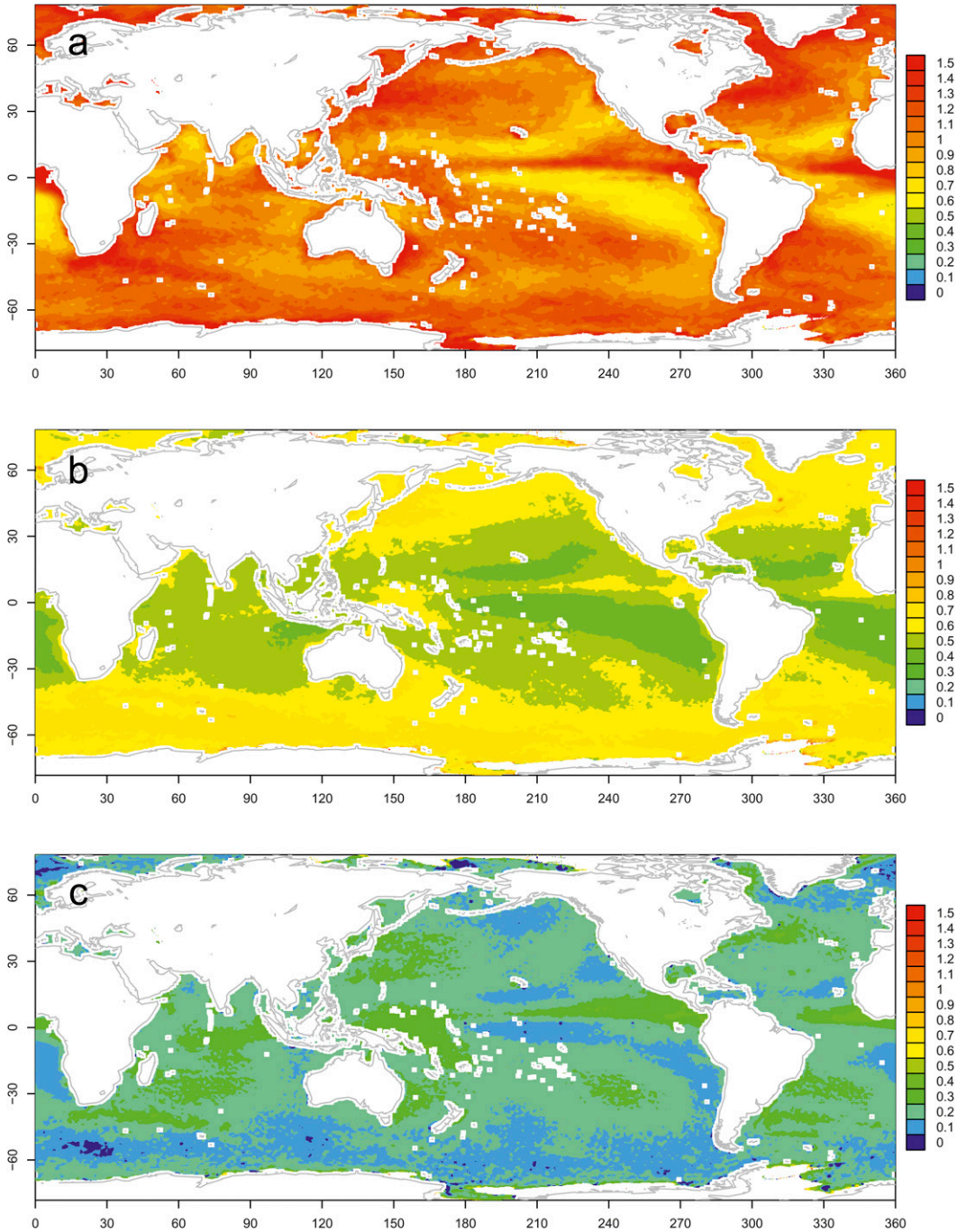


FIG. 3. Maps of estimated error standard deviations based on all platforms for 2004. The estimated error standard deviations are for the (a) background, (b) observations, and (c) analysis.

estimates of analysis errors for each location for each synoptic time, based on the relationships discovered using the DD between errors and other variables, including location, wind speed, density of observations, etc. Fourth, we will use the results to tune the VAM and iterate the process.

*Acknowledgments.* This work was supported by grants made through the NASA Making Earth Science Data Records for Use in Research Environments (MEaSUREs) program and through NASA ROSES 2011 [Grant NNX11AO25A to Remote Sensing Systems (RSS), Santa Rosa, California].

# APPENDIX

## Error Variances and the DD

When $i = j$, it is easy to validate that $\overline{C_x}$ is $\overline{\varepsilon_x^2}$, the error variance of $x$. Here $\varepsilon_x$, the error of $x$, is defined by $x = T + \varepsilon_x$, where $T$ is the truth. Each difference term in Eqs. (1)–(3) can be written in terms of $\varepsilon$, since, for example, $A - B = (T + \varepsilon_A) - (T + \varepsilon_B) = \varepsilon_A - \varepsilon_B$. Then, rearranging terms and taking sample averages of Eqs. (1)–(3), we obtain each $\overline{C_x}$ as the sum of four error covariance terms. For example, Eq. (1) becomes

$$\overline{C_B} = \overline{\varepsilon_A \varepsilon_O} - \overline{\varepsilon_A \varepsilon_B} - \overline{\varepsilon_B \varepsilon_O} + \overline{\varepsilon_B^2}. \quad \text{(A1)}$$

By assumption we know the error covariances. First, $O$ and $B$ are assumed to be independent, so $\overline{\varepsilon_O \varepsilon_B} = 0$. Here and the rest of this paragraph, we first think of the overbar as an expectation operator and at the end make the assumption that we may replace expectation with sample average. Second, since $A$ is optimal, its errors must be uncorrelated with $O - B$, $\overline{\varepsilon_A (O - B)} = 0$ or $\overline{\varepsilon_A \varepsilon_O} - \overline{\varepsilon_A \varepsilon_B} = 0$. Therefore, in Eq. (A1) the first two terms together are zero, the third term is zero, leaving the expected result, $\overline{C_B} = \overline{\varepsilon_B^2}$. Very similar steps lead to the analogous result for $C_O$. For $C_A$, an additional relationship, $\overline{\varepsilon_A^2} - \overline{\varepsilon_A \varepsilon_B} = 0$, is used. This follows since $A - B$ is a linear combination of the $O - B$, and the errors of $A$ are uncorrelated with $O - B$. See Fig. 1 in Desroziers et al. (2005) and the related discussion for a geometric interpretation.

Also, when $i = j$, it is easy to see that Eqs. (1)–(3) imply $C_O + C_B = (O - B)^2$, $C_O - C_A = (O - A)^2$, and $C_B - C_A = (A - B)^2$. These are identities and may be averaged over any sample. Solving for the $C_x$ and averaging over the sample gives

$$\overline{C_B} = \frac{1}{2}\left[\overline{(O-B)^2} + \overline{(A-B)^2} - \overline{(O-A)^2}\right], \quad \text{(A2)}$$

$$\overline{C_O} = \frac{1}{2}\left[\overline{(O-B)^2} + \overline{(O-A)^2} - \overline{(A-B)^2}\right], \quad \text{and} \quad \text{(A3)}$$

$$\overline{C_A} = -\frac{1}{2}\left[\overline{(O-A)^2} + \overline{(A-B)^2} - \overline{(O-B)^2}\right]. \quad \text{(A4)}$$

## REFERENCES


Atlas, R., R. N. Hoffman, J. Ardizzone, S. M. Leidner, J. C. Jusem, D. K. Smith, and D. Gombos, 2011: A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bull. Amer. Meteor. Soc.,* **92,** 157–174.

Bormann, N., and P. Bauer, 2010: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data. *Quart. J. Roy. Meteor. Soc.,* **136,** 1036–1050.

——, A. Collard, and P. Bauer, 2010: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data. *Quart. J. Roy. Meteor. Soc.,* **136,** 1051–1063.

——, ——, and ——, 2011: Observation errors and their correlations for satellite radiances. *ECMWF Newsletter,* No. 128, ECMWF, Reading, United Kingdom, 17–22.

Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand, 2006: Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Quart. J. Roy. Meteor. Soc.,* **132,** 543–565.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.,* **131,** 3385–3396, doi:10.1256/qj.05.108.

Gierach, M. M., T. Lee, D. Turk, and M. McPhaden, 2012: Biological response to the 1997–98 and 2009–10 El Niño events in the equatorial Pacific Ocean. *Geophys. Res. Lett.,* **39,** L10602, doi:10.1029/2012GL051103.

Hoffman, R. N., and S. M. Leidner, 2010: Some characteristics of time interpolation errors for fluid flows. *J. Atmos. Oceanic Technol.,* **27,** 1255–1262.

——, ——, J. M. Henderson, R. Atlas, J. V. Ardizzone, and S. C. Bloom, 2003: A two-dimensional variational analysis method for NSCAT ambiguity removal: Methodology, sensitivity, and tuning. *J. Atmos. Oceanic Technol.,* **20,** 585–605.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 523–533.

Lupu, C., P. Gauthier, and S. Laroche, 2012: Assessment of the impact of observations on analyses derived from observing system experiments. *Mon. Wea. Rev.,* **140,** 245–257.

McCarthy, G., and Coauthors, 2012: Observed interannual variability of the Atlantic meridional overturning circulation at 26.5°N. *Geophys. Res. Lett.,* **39,** L19609, doi:10.1029/2012GL052933.

Miyoshi, T., E. Kalnay, and H. Li, 2013: Estimating and including observation-error correlations in data assimilation. *Inverse Probl. Sci. Eng.,* **21,** 387–398, doi:10.1080/17415977.2012.712527.

Portabella, M., and A. Stoffelen, 2009: On scatterometer ocean stress. *J. Atmos. Oceanic Technol.,* **26,** 368–382.

Stoffelen, A., and D. Anderson, 1997: Ambiguity removal and assimilation of scatterometer data. *Quart. J. Roy. Meteor. Soc.,* **123,** 491–518.

Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.,* **131,** 2961–3012.

Wahba, G., and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.,* **108,** 1122–1143.