THE SPATIAL STRUCTURE OF OBSERVATION ERRORS IN ATMOSPHERIC MOTION VECTORS

Niels Bormann, Sami Saarinen, Jean-Noël Thépaut, and Graeme Kelly

European Centre for Medium-Range Weather Forecasts (ECMWF)
Shinfield Park, Reading, Berkshire RG2 9AX, UK, n.bormann@ecmwf.int

ABSTRACT

This study investigates in detail the spatial correlations of random errors in Atmospheric Motion Vectors (AMVs). A good specification of the observation error is essential to assimilate any kind of observation for Numerical Weather Prediction in a near-optimal way. For AMVs, height assignment, quality control procedures, and other factors introduce spatially correlated errors.

The spatial structure of the error correlations is investigated based on a one-year dataset of pairs of collocations between AMVs and radiosonde observations. Assuming spatially uncorrelated sonde errors, the spatial AMV error correlations are obtained over dense sonde networks. Results for operational IR and WV wind datasets from METEOSAT-5 and 7, GOES-8 and 10, and GMS-5 are presented.

Winds from all five datasets show statistically significant spatial error correlations for distances up to about 800 km, with little difference between satellites, channels, or vertical levels. Broader correlations are found for tropical regions. The correlations exhibit anisotropic structures with, for instance, longer correlation scales in South-North direction for the v-wind component. The study estimates the spatially correlated part of the annual mean AMV wind component error for high-level Northern Hemisphere winds as about 2.7–3.5 m/s. The findings have a number of important implications for the use of AMVs in data assimilation.

1 Introduction

This study characterises statistically the spatial structure of observation errors in Atmospheric Motion Vectors (AMVs) by analysing a large dataset of pairs of AMV/radiosonde collocations. AMVs from geostationary satellite data provide excellent temporal and spatial coverage and therefore are an important input to data assimilation systems (e.g., Bouttier and Kelly 2001). Most wind producers now provide datasets at 160 km resolution or higher (e.g., Nieman et al. 1997, Schmetz et al. 1993).

A good specification of the random and systematic errors of any observation is essential in order to extract information from the observation in an assimilation system in a near optimal way. The errors assigned to the observations together with an estimate for the error in the First Guess fields determine the weighting of both in the analysis system and therefore which features are assimilated from the observations (e.g., Daley 1993). For technical reasons, error specifications are frequently simplified. For instance, First Guess errors are assumed to be isotropic and observation errors are assumed to be uncorrelated between neighbouring observations (e.g., Derber and Bouttier 1999, Daley 1993).

Many authors have suggested that AMVs possess spatially correlated errors (e.g., Rohn et al. 2001, Butterworth and Ingleby 2000) and thus invalidate the assumption on uncorrelated observation errors. For instance, the height assignment is usually aided by forecast temperature profiles which inherently have spatially correlated errors (e.g., Daley 1993). Also, tracking of similar cloud structures in neighbouring segments may lead to similar tracking or height assignment errors. Furthermore, quality control procedures tend to favour winds which are consistent with their neighbours (e.g., Holmlund 1998), therefore enhancing the chance of correlated errors. As current data assimilation systems do not account for such correlated errors, AMVs are frequently thinned to a lower resolution (e.g., $1.25° \approx 140$ km, Rohn et al. 2001), or observation errors are inflated, to avoid overfitting. This contribution characterises the error correlations and thus provides guidance for a refined use of AMVs in data assimilation systems. A more detailed report is provided in Bormann et al. (2002).

2 Method and Data

2.1 Overview

The calculations presented in this study use a large number of pairs of collocations between an AMV and a radiosonde, covering one year. The method to derive spatial error correlations from this database is similar to the one used for short-range forecast errors (e.g., Daley 1993, Hollingsworth and Lönnberg 1986, Rutherford 1972), and it is based on the assumption that observation errors from sondes are spatially uncorrelated. Therefore, any correlation between the AMV-sonde differences of two stations are attributed to spatially correlated AMV errors. Grouping the collocation pairs from a dense sonde network by station separation allows a characterisation of the average spatial structure of the AMV error correlations over this network. This can be done either in an isotropic way (grouping by station distance) or anisotropically (grouping by S–N and W–E separation). Note that the smallest distance between two sonde stations is typically around 150 km, so the correlation structure for smaller distances cannot be resolved. The method has been employed numerous times for forecast data or climatologies, and a more detailed discussion can be found in Daley (1993).

The magnitude of the spatially correlated AMV error can be estimated by extrapolating the separation/ correlation relationship to zero separation. A suitable correlation function is used to do this, as is further discussed below. The extrapolated correlation at zero separation partitions the variance of the AMV-sonde differences into spatially correlated and spatially uncorrelated parts. The former is the spatially correlated part of the AMV observation error, whereas the latter is made up of the following components: the spatially uncorrelated AMV error, the error for the sonde observations, and any errors arising from the mismatch of representativeness between sondes (point measurements) and AMVs (area averages).

2.2 Collocations

To derive the statistics presented here we use the AMV datasets summarised in Table 1 for the period 1 January–31 December 2001, collocated with wind observations from all available radiosonde or pilot reports[1] (e.g., Fig. 1). Poor AMV or sonde observations have been eliminated by requiring a quality indicator of 60% or more for the EUMETSAT data (e.g., Rohn et al. 2001, Holmlund 1998, quality indicators were not yet available for the other wind datasets), and by rejecting any collocations with more than 18 m/s vector difference (approximately three times the

---

[1]For simplicity, radiosondes and pilot reports will be referred to as "sondes" in the remainder of the text.

Table 1: AMV data used.

Satellite	Producer	Data used
GOES-8, GOES-10	NOAA/NESDIS	operational IR and WV winds
MET-5, MET-7	EUMETSAT	operational IR and WV cloud track winds from 160 km segments
GMS-5	JMA	operational IR and WV winds

standard deviation of the departures). Note that for technical reasons GOES and GMS WV winds may include AMVs from clear-sky areas, whereas METEOSAT (MET) WV winds are cloud track winds only.

We collocate one AMV and one sonde observation if they are less than 150 km apart, have less than 25 hPa separation in the vertical, and are separated less than 1.5 h in time. These criteria follow CGMS recommendations (e.g., Velden and Holmlund 1998). Further experimentation has revealed no significant sensitivity of our results to tightening the spatial collocation criterion. Each AMV/sonde collocation is matched up with successful collocations from all other sonde stations if the AMVs originate from the same imagery, and the difference between the assigned pressures is less than 150 hPa. The latter criteria restrict our study to spatial correlations. Biases are removed for each station by subtracting the mean difference vector for the vertical layer considered in the statistics.

The method can be used only in regions with dense high-quality sonde networks (e.g., North America, Europe, South-East Asia, or Australia/New Zealand; Fig. 1) to achieve suitable sample sizes. As a result, statistics can mainly be derived over land only. Furthermore, we derived statistics for the entire year over larger regions (e.g., Northern Hemisphere[2] extra-tropical part of the respective satellite disk) and assume uniform errors and error correlations for these regions.
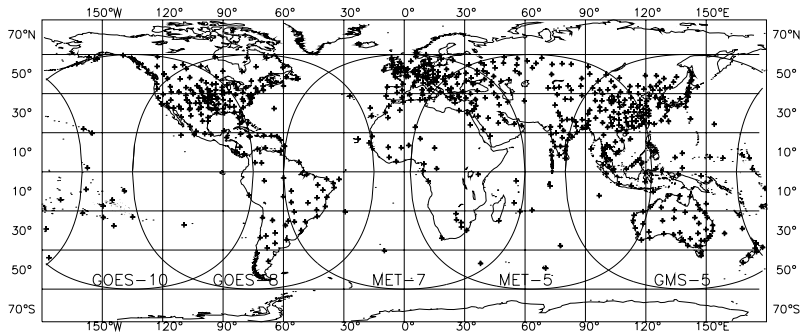

Figure 1: Map of radiosonde or pilot sites considered in this study. Also shown are the outlines of the disks viewed by each geostationary satellite.

2.3   Correlation function

For the isotropic component we derived a least squares fit of a correlation function to our empirical correlation data. We found the isotropic assumption a good approximation for the $\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle$ correlations, and will therefore focus the discussion of the isotropic part on this quantity (Δu and Δv denote the difference between the collocated AMV and radiosonde in the u and v

[2]Geographical areas are referred to as follows: Northern Hemisphere (NH): North of 20° N; Tropics: 20° N–20° S; Southern Hemisphere (SH): South of 20° S.

component, respectively). The correlation function is primarily used to extrapolate the correlation data in a statistically reasonable way to zero separation to estimate the AMV errors. We use the following function R of the station distance r:

$$R(r) \;=\; R_0 \left(1 + \frac{r}{L}\right) e^{-\frac{r}{L}} \tag{1}$$

with the intercept $R_0 > 0$ and the length scale $L > 0$ as fitting parameters. This correlation function has been used by a number of other authors (e.g., Daley 1993, Thiébaux 1985). It is chosen here as it usually fits our correlation data within estimated 95 % confidence intervals. The fits were calculated from data grouped into 100 km bins with weights of each data point determined by 95 % confidence intervals for the correlations. The confidence intervals were estimated through a t-test, based on sub-sampled data.

3 Results

3.1 Isotropic error correlations

We will first present results for the isotropic part of the $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ departure correlations for the different satellites and geographical regions. Figure 2 shows the Northern Hemisphere correlations between the AMV-sonde differences as a function of station distance for WV winds from all levels for the five satellites.
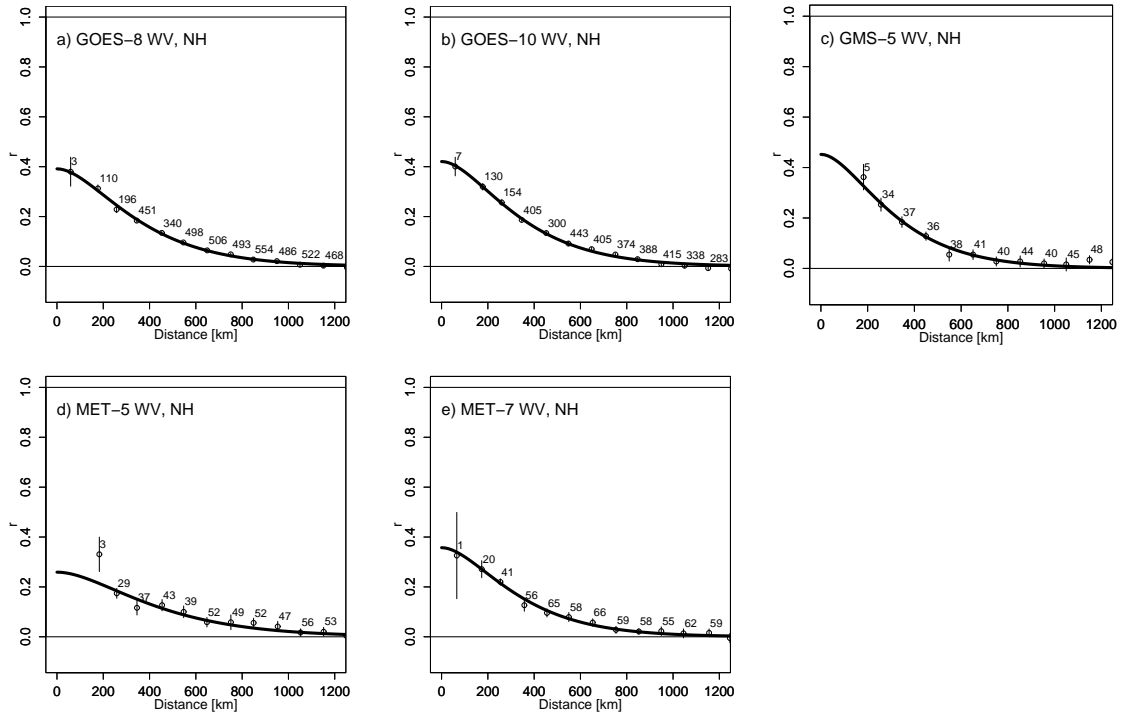

Figure 2: a) $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ AMV-sonde departure correlations for Northern Hemisphere GOES-8 WV winds from all levels as a function of station separation. Error bars indicate 95 % confidence intervals calculated as described in the main text. Also shown is the fitted correlation function and the number of collocations used per data point (in hundreds). b) As a), but for GOES-10 WV winds. c) As a), but for GMS-5 WV winds. d) As a), but for MET-5 WV cloud track winds. e) As a), but for MET-7 WV cloud track winds.

There are statistically significant correlations in the AMV-sonde differences for distances up to about 800 km. As expected, the correlations are close to zero for station distances larger than about 1000 km. Note that Fig. 2 shows the correlations between the AMV-sonde differences; to obtain an estimate of the AMV *error* correlations we would need to normalise the values to 1 at zero separation.

The departure correlations show no statistically significant differences for the five WV AMV datasets, except for Northern Hemisphere MET-5 winds which show significantly broader and flatter departure correlations. The reasons for the different behaviour of MET-5 WV winds are not well understood, but might be related to the East Asian sonde network and geography rather than the AMV data. The lack of difference between the four other datasets is somewhat counterintuitive, given that the AMVs originate from three producers with considerably different processing, particularly in terms of quality control. However, the datasets are evaluated over different regions, so similarities or differences can be a combined result of different processing and geographical differences in the AMV errors.

Winds from the IR channels of the five satellites show similar results (not shown). Note that fewer collocations are available for IR winds as these cover a larger vertical range. The results are therefore less accurate. Error correlation structures derived for different vertical layers for IR winds also do not differ significantly, in the sense that the correlation scales of the fitted correlation function are usually not significantly different (not shown).
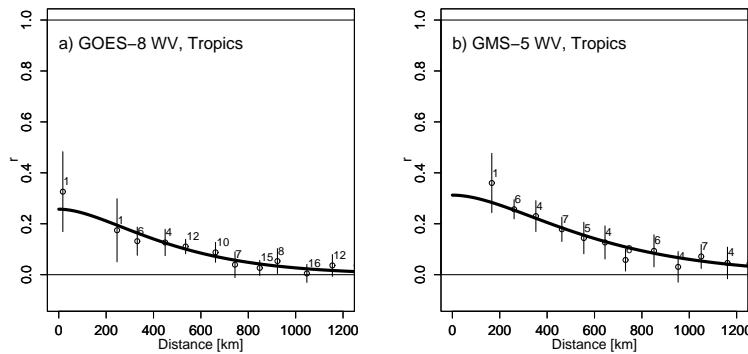

Figure 3: a) As Fig. 2a, but for tropical GOES-8 WV winds. b) As a), but for GMS-5 WV winds.

The departure correlations tend to be flatter and broader over the Tropics for most winds (cf, Figures 2 and 3). These differences in the correlation scales are mostly statistically significant, even though results for the Tropics are less accurate. Over the Southern Hemisphere, meaningful results could only be obtained for GMS-5 WV winds, and these are not significantly different from their Northern Hemisphere counterparts (not shown).

3.2   Estimation of AMV errors

We will now estimate the spatially correlated part of the observation error in AMVs by extrapolating to zero distance the distance/correlation relationships using the fitted correlation functions. We use the $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ departure correlations and assume that the errors in each wind component are the same.

Our statistics give spatially correlated AMV errors for WV wind components in the Northern Hemisphere of 3.1–3.5 m/s, except for MET-5, where the problems noted earlier produce suspiciously low error estimates of 2.6 m/s (Table 2). The differences between the other four wind datasets are not statistically significant. Also, within the accuracy of the method used, errors for

Table 2: Estimates of the AMV wind component error σ_{AMV} [m/s], including uncertainty estimates. The values are derived for AMVs from all levels. "n/a" indicates insufficient data for reliable statistics; an entry in angle brackets indicates a suspect estimate.

Data	WV		IR	
	σ_{AMV} [m/s] NH	σ_{AMV} [m/s] Tropics	σ_{AMV} [m/s] NH	σ_{AMV} [m/s] Tropics
GOES-8	3.3(1)	2.3(5)	3.3(1)	2.2(4)
GOES-10	3.5(1)	n/a	3.1(1)	n/a
MET-5	2.6(3)	[1.6(5)]	2.7(6)	2.4(6)
MET-7	3.1(3)	n/a	2.8(3)	n/a
GMS-5	3.5(4)	2.5(5)	2.7(5)	n/a

the Southern Hemisphere GMS-5 WV winds show similar values (not shown). The errors for the tropical winds tend to be significantly lower, with values of 1.6–2.5 m/s. IR winds give similar, but slightly smaller errors with overall Northern Hemisphere errors of around 2.7–3.3 m/s for all levels.

For high and medium levels, the estimates for the correlated part of the AMV wind component error agree well with independent estimates for the total AMV component error based on departures from the ECMWF First Guess together with estimates for the First Guess errors (not shown). This gives some indication that the spatially uncorrelated part of AMV errors is likely to be small, and the spatially correlated part dominates. For higher levels, our new estimates for the annual mean AMV error are considerably lower than AMV errors used at most NWP centres (e.g., Tsuyuki 2000). Compared to the wind component errors assigned to radiosondes in the ECMWF system, the error estimates for the satellite winds presented here are larger by about 20–30 %.

3.3   Anisotropic error correlations

Meaningful statistics for the anisotropic structure of the error correlations could only be produced for Northern Hemisphere WV winds, and Northern Hemisphere IR winds from GOES-8 and GOES-10. These reveal a considerably anisotropic structure of the error correlations. For example, the S–N correlation scales for the $\langle \Delta v, \Delta v \rangle$ departure correlations are much broader than the W–E correlation scales for most WV or IR winds considered (e.g., Fig. 4), except for GMS-5 WV winds (not shown). Apart from a slightly more diagonal orientation, the correlation structures shown in Fig. 4 show similarities with error correlations typical for short-term forecast errors (e.g., Daley 1993, Hollingsworth and Lönnberg 1986).

4   Discussion and Conclusions

We have characterised in detail the spatial structure of errors in AMVs by analysing 12 months of pairs of collocations of AMVs and radiosondes and assuming spatially uncorrelated errors in the sonde observations. The analysis provides estimates of the spatially correlated part of the AMV component error. The main findings are:

- AMVs from five different geostationary satellites and three different winds producers all show statistically significant spatial error correlations on scales up to about 800 km. The
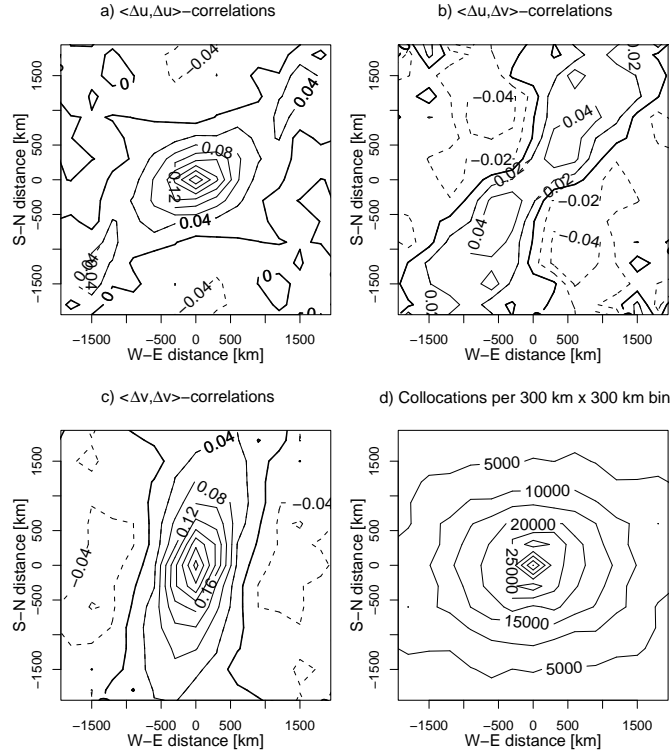
Figure 4: a) Spatial structure of the $\langle \Delta u, \Delta u \rangle$ AMV-sonde departure correlations for Northern Hemisphere GOES-8 WV winds as a function of W–E and S–N distance from an AMV located at (0 km,0 km). Note that only the upper half of the plot was computed from the available data; the lower half was obtained through point symmetry. The binning box size is 300 km × 300 km. b) As a), but for the $\langle \Delta u, \Delta v \rangle$ departure correlations. c) As a), but for the $\langle \Delta v, \Delta v \rangle$ departure correlations. d) The number of collocations per 300×300 km bin.

  correlations are similar for winds from different satellites, spectral channels, or vertical levels. Tropical error correlations tend to be broader than midlatitude ones.

- The spatially correlated part of the annual mean AMV component error is about 3.1–3.5 m/s for extra-tropical WV winds above 400 hPa and 2.7–3.3 m/s for IR winds, depending on satellite. The estimates have a relatively large uncertainty (0.1-0.5 m/s), and within this accuracy no differences in the AMV errors from different satellites or producers can be reported. The correlated AMV errors for the tropics tend to be smaller (2.2–2.5 m/s).

- The AMV-sonde departure correlations show considerable anisotropy. For instance, $\langle \Delta v, \Delta v \rangle$-departure correlations are broader in S–N direction than in W–E direction for most winds.

The similarities found for the spatial error correlations for different wind datasets are striking, particularly given the differences in the processing and quality control between different winds producers. This confirms that spatial error correlations are inherent in the AMV approach, and it is unlikely that such correlations can be removed in the winds derivation. The use of temperature forecasts for the height assignment is likely to be the largest contribution to these correlated errors. This is supported by the fact that the correlations found in this study share some of the characteristics of correlations in short term forecast errors. For instance, the length scales for the AMV error correlations compare favourably with length scales in temperature forecast errors (e.g., Derber and Bouttier 1999). Considering the entire AMV dataset, the correlated observation errors imply that the larger-scale spatial structures represented in the AMV dataset have larger errors than the small-scale structures (e.g., Bormann et al. 2002).

The findings about the spatial correlations of the AMV errors highlight some shortcomings in the use of AMVs at many data assimilation centres. The satellite winds indeed invalidate the assumption of uncorrelated observation errors inherent in many data assimilation systems. Furthermore, the correlation scales found in this study are much larger than the thinning scales typically applied to AMVs in an attempt to suppress the impact of spatially correlated errors (1.25–$2.5° \approx$ 140–275 km, e.g., Rohn et al. 2001). However, even a suboptimal specification of the error covariances can be used to extract some information from observations, even though the reduction in the analysis error will not be optimal. This has been shown in theoretical studies and by practical experience with positive forecast impact from the assimilation of AMVs with such suboptimal settings (e.g., Bouttier and Kelly 2001, Daley 1993).

The discrepancies between our results and current assimilation practise suggest that there is some scope for improvement in the assimilation of AMVs through a revision of the assumed error characteristics, based on correlated errors, revised thinning or error inflation. More realistic spatially correlated observation errors in a data assimilation system will alter the filtering properties of the system compared to uncorrelated observation errors, and further investigations are necessary to characterise the influence and relevance of the AMV error correlations in data assimilation.

Acknowledgements

References

Bormann, N., S. Saarinen, G. Kelly, and J.-N. Thépaut, 2002: The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. EUMETSAT/ECMWF Fellowship Programme Research Report 12, ECMWF, Reading, U.K. [submitted to *Mon. Wea. Rev.*].

Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4-DVAR assimilation system. *Quart. J. Roy. Meteor. Soc.*, **127**, 1496–1488.

Butterworth, P., and N. B. Ingleby, 2000: Recent developments in the use of satellite winds at the UK Met. Office. In *Proceedings of the Fifth International Winds Workshop*, Lorne, Australia, EUMETSAT, 151–159.

Daley, R., 1993: *Atmospheric data analysis.* Cambridge University Press, Cambridge, UK, 460 pp.

Derber, J., and F. Bouttier, 1999: A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus*, **51A**, 195–221.

Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.

Holmlund, K., 1998: The utilization of statistical properties of satellite-derived atmospheric motion vectors to derive quality indicators. *Wea. Forecasting*, **13**, 1093–1104.

Nieman, S. J., W. P. Menzel, C. M. Hayden, D. Gray, S. T. Wanzong, C. S. Velden, and J. Daniels, 1997: Fully automated cloud-drift winds in NESDIS operations. *Bull. Amer. Meteor. Soc.*, **78**, 1121–1133.

Rohn, M., G. Kelly, and R. W. Saunders, 2001: Impact of a new cloud motion wind product from Meteosat on NWP analyses. *Mon. Wea. Rev.*, **129**, 2392–2403.

Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.*, **29**, 809–815.

Schmetz, J., K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and L. Van De Berg, 1993: Operational cloud-motion winds from Meteosat infrared images. *J. Appl. Meteor.*, **32**, 1206–1225.

Thiébaux, H. J., 1985: On approximations to geopotential and wind-field correlation structures. *Tellus*, **37A**, 126–131.

Tsuyuki, T., 2000: Summary of comments from NWP centres represented in WGNE on the large differences in satellite wind observation errors assigned at NWP centers. Report (available under http://www.met-office.gov.uk/research/interproj/nwpsaf/satwind_report/cgms_comments_015.pdf), CGMS, Geneva, Switzerland.

Velden, C. S., and K. Holmlund, 1998: Report from the working group on verification and quality indices (WG III). In *Proceedings of the Fourth International Winds Workshop*, Saanenmöser, Switzerland, EUMETSAT, 19–20.