EUMETSAT/ECMWF Fellowship Programme,
Research Report No. 12

The spatial structure of observation errors in Atmospheric Motion Vectors from geostationary satellite data

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

Submitted to Mon. Wea. Rev.

February 2002

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading
RG2 9AX
library@ecmwf.int

Series: EUMETSAT/ECMWF Fellowship Program - Research Reports

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/pressroom/publications/

Abstract

This study investigates and quantifies in detail the spatial correlations of random errors in Atmospheric Motion Vectors (AMVs) derived by tracking structures in imagery from geostationary satellites. 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 or quality control procedures 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. Even broader correlations are found for tropical regions. The correlations exhibit considerable anisotropic structures with, for instance, longer correlation scales in South-North direction for the v-wind component, and are comparable to error correlations for short-term forecasts. 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. Some seasonal variation is found for these errors with larger values in winter. The findings have a number of important implications for the use of AMVs in data assimilation.

1 Introduction

This paper 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 most global and some mesoscale data assimilation systems (e.g., Bouttier and Kelly 2001, Soden et al. 2001). These "satellite winds" are currently derived operationally at EUMETSAT, NOAA/NESDIS or JMA from five satellites (METEOSAT "MET" 5 & 7, GOES 8 & 10, GMS 5) by tracking clouds in the infrared (IR), water-vapour (WV), or visible (VIS) channel, or clear-sky features in the WV channel. The spatial coverage has increased dramatically in recent years, with most wind producers now providing datasets at 160 km resolution or higher (e.g., Nieman et al. 1997, Schmetz et al. 1993). Automatic quality control procedures have been developed, and winds producers now endeavour to provide quality flags which summarise the consistency of the derived wind vectors (e.g., Holmlund et al. 2001, Holmlund 1998).

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., Lorenc et al. 2000, 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. A number of aspects of the winds derivation is responsible for this. 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., $1.25^\circ \approx 140$ km, Rohn et al. 2001), or observation errors are inflated, to avoid overfitting. Given the uncertainty about AMV errors and their spatial structure, it is not surprising that observation errors assigned to AMVs at different

Numerical Weather Prediction (NWP) centres vary by more than a factor of 2 (e.g., Tsuyuki 2000).

This paper characterises and quantifies in detail the spatial correlations of AMV errors, gives model-independent estimates of AMV errors, and thus provides important guidance for the use of AMVs in data assimilation systems. The results are an important first step towards the use of spatially correlated observation errors, or the careful refinement of thinning schemes or observation errors used for AMVs.

The structure of the paper is as follows. We first discuss how AMVs and other wind observations are used in this study to characterise the spatial error correlation structure and to estimate the spatially correlated part of the error in the satellite winds. We then present our results on the error correlations and estimates of the AMV observation error. These will be discussed in the last section in which a summary of the main conclusions is also provided.

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. For each pair, wind observations from two sonde stations have been collocated with a different satellite wind derived from the same imagery. 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).

We use the method to characterise anisotropic as well as isotropic correlation structures. We limit quantitative satellite intercomparisons to isotropic structures due to sample limitations. We found the isotropic assumption most valid 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 component, respectively). As pointed out by Hollingsworth and Lönnberg (1986), this quantity is invariant under coordinate transforms.

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). A more detailed discussion of the contributions to the correlations and the assumptions inherent in our approach can be found in Appendix A.

Table 1: AMV data used.

Satellite	Sub-satellite point	Producer	Data used
GOES-8	75° W	NOAA/NESDIS	operational IR and WV winds
GOES-10	135° W	NOAA/NESDIS	as GOES-10
MET-5	63° E	EUMETSAT	operational IR and WV cloud track winds from 160 km segments, with quality indicator
MET-7	0° E	EUMETSAT	as MET-5
GMS-5	140° E	JMA	operational IR and WV winds

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 standard deviation of the departures). The latter criterion is similar to the First Guess check applied in the ECMWF system (e.g., Järvinen and Unden 1997). Note that for technical reasons GOES and GMS WV winds may include AMVs from clear-sky areas, whereas METEOSAT WV winds are cloud track winds only.

The collocation criteria are as follows: 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) and have been found a useful compromise between too large spatial separations (which limit the accuracy of the spatial characterisation of

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

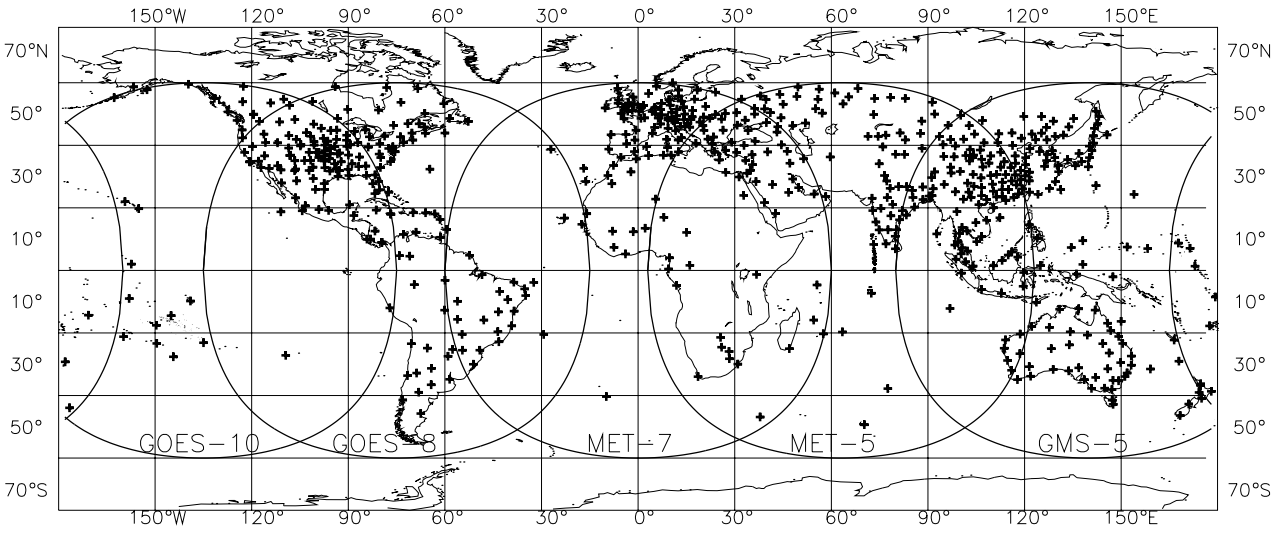

Figure 1: Map of radiosonde or pilot sites considered in this study. Also shown are the outlines of the disks viewed by each geostationary satellite.

the error structure) and too tight criteria (which limit the sample size). Further experimentation has revealed no significant sensitivity of our results to tightening the spatial collocation criterion. Care has to be taken that each AMV and each sonde observation is collocated only once to avoid introducing spurious correlations. 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. Only stations with more than ten collocations are used to facilitate a reasonable bias removal.

It should be pointed out here that the method is very data intensive and requires a good sample size for useful statistics. It can thus be used only in regions with dense high-quality sonde networks, such as North America, Europe, South-East Asia, or Australia/New Zealand (Fig. 1). As a result, statistics can mainly be derived over land only, whereas in data assimilation systems the satellite winds have an important impact over sea and are frequently partially excluded over land. Also, the need of *pairs* of AMV/sonde collocations within a pressure range further limits the available sample size. To achieve suitable sample sizes, we derived statistics for the entire year over larger regions (e.g., Northern Hemisphere extra-tropical part of the respective satellite disk) and assume uniform errors and correlations for these regions. Seasonal variations are investigated for GOES winds only.

2.3 Correlation function

To quantify our results for the isotropic component we derived a least squares fit of a correlation function to our empirical correlation data. The purpose of fitting a correlation function is twofold: 1) The function is used to extrapolate the correlation data in a statistically reasonable way to zero separation to estimate the AMV errors. 2) It allows a better quantitative comparison of the results for different satellites and areas. We tried a range of suitable correlation functions (e.g., Gaussian, Bessel series; Thiébaux 1976), and decided on the following function R of the station distance r:

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

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. L represents the length scale of R in the usual way (Daley 1993)

$$L^2 = -\frac{2\,R(r)}{\nabla^2 R(r)}|_{r=0} \tag{2}$$

Note, however, that $R(L) = R_0\,\frac{2}{e}$. It should also be pointed out that R is positive definite for all $L > 0$ in $\mathbf{R}^2$ only, whereas for $\mathbf{R}^3$ or the 2-dimensional sphere this may not be the case for some parameter choices (e.g., Weber and Talkner 1993). We found this not to be a problem for our application.

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. We found the results to be insensitive to reasonable choices of the bin size. Estimating the confidence in our calculations is not straightforward as the samples are inherently not all independent. We thus obtained uncertainty estimates in the following way: For each station separation "bin", we randomly divided the data into 10 similarly sized subsamples, and calculated separate correlations for each subsample. We then derived confidence intervals from these sub-sampled estimates using the t-test. Confidence intervals for the fitting parameters of the correlation function were also calculated from these 10 subsamples.

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[2] correlations between the AMV-sonde differences as a function of station distance for WV winds from all levels for the five satellites. The fitted correlation function is plotted as well, with the fitting parameters given in Table 2.

There are statistically significant correlations in the AMV-sonde differences for distances up to about 800 km,

[2]In the following, 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.

Table 2: Fitting parameters R_0 and L for the isotropic part of the $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ correlations vs station distance for WV winds. Also given is an estimate for the correlated part of the AMV error (σ_{AMV}) and estimates for 95 % confidence intervals. We only show values for regions for which meaningful results could be obtained, using data for the whole year 2001.

Satellite	R_0	L [km]	σ_{AMV} [m/s]
NH: all levels			
GOES-8	$0.39^{+0.02}_{-0.02}$	198^{+6}_{-6}	$3.3^{+0.1}_{-0.1}$
GOES-10	$0.42^{+0.02}_{-0.02}$	190^{+5}_{-5}	$3.5^{+0.1}_{-0.1}$
MET-5	$0.26^{+0.06}_{-0.03}$	242^{+40}_{-40}	$2.6^{+0.3}_{-0.2}$
MET-7	$0.35^{+0.05}_{-0.04}$	185^{+19}_{-15}	$3.1^{+0.3}_{-0.2}$
GMS-5	$0.45^{+0.11}_{-0.05}$	176^{+21}_{-20}	$3.5^{+0.5}_{-0.2}$
NH: above 400 hPa			
GOES-8	$0.38^{+0.02}_{-0.02}$	196^{+4}_{-4}	$3.3^{+0.1}_{-0.1}$
GOES-10	$0.41^{+0.02}_{-0.02}$	191^{+6}_{-6}	$3.4^{+0.1}_{-0.1}$
MET-5	$0.24^{+0.04}_{-0.04}$	260^{+40}_{-20}	$2.5^{+0.5}_{-0.4}$
MET-7	$0.35^{+0.08}_{-0.05}$	185^{+27}_{-21}	$3.1^{+0.4}_{-0.2}$
GMS-5	$0.46^{+0.07}_{-0.05}$	169^{+20}_{-15}	$3.5^{+0.3}_{-0.2}$
NH: below 400 hPa			
GOES-8	$0.49^{+0.05}_{-0.04}$	168^{+10}_{-8}	$4.0^{+0.2}_{-0.2}$
GOES-10	$0.47^{+0.09}_{-0.05}$	184^{+19}_{-16}	$3.9^{+0.4}_{-0.2}$
MET-7	$0.40^{+0.11}_{-0.10}$	150^{+40}_{-20}	$3.4^{+0.5}_{-0.4}$
Tropics: all levels			
GOES-8	$0.25^{+0.20}_{-0.05}$	270^{+110}_{-70}	$2.3^{+0.7}_{-0.2}$
MET-5	$0.13^{+0.23}_{-0.04}$	360^{+180}_{-130}	$1.6^{+0.7}_{-0.2}$
GMS-5	$0.30^{+0.17}_{-0.06}$	350^{+150}_{-90}	$2.5^{+0.7}_{-0.3}$
Tropics: above 400 hPa			
GOES-8	$0.25^{+0.08}_{-0.06}$	260^{+80}_{-40}	$2.3^{+0.4}_{-0.3}$
MET-5	$0.13^{+0.12}_{-0.05}$	370^{+200}_{-130}	$1.6^{+0.6}_{-0.3}$
GMS-5	$0.30^{+0.13}_{-0.05}$	350^{+190}_{-60}	$2.5^{+0.5}_{-0.3}$
SH: all levels			
GMS-5	$0.48^{+0.13}_{-0.11}$	160^{+100}_{-30}	$3.6^{+0.5}_{-0.4}$
SH: above 400 hPa			
GMS-5	$0.46^{+0.13}_{-0.13}$	150^{+110}_{-20}	$3.5^{+0.5}_{-0.6}$

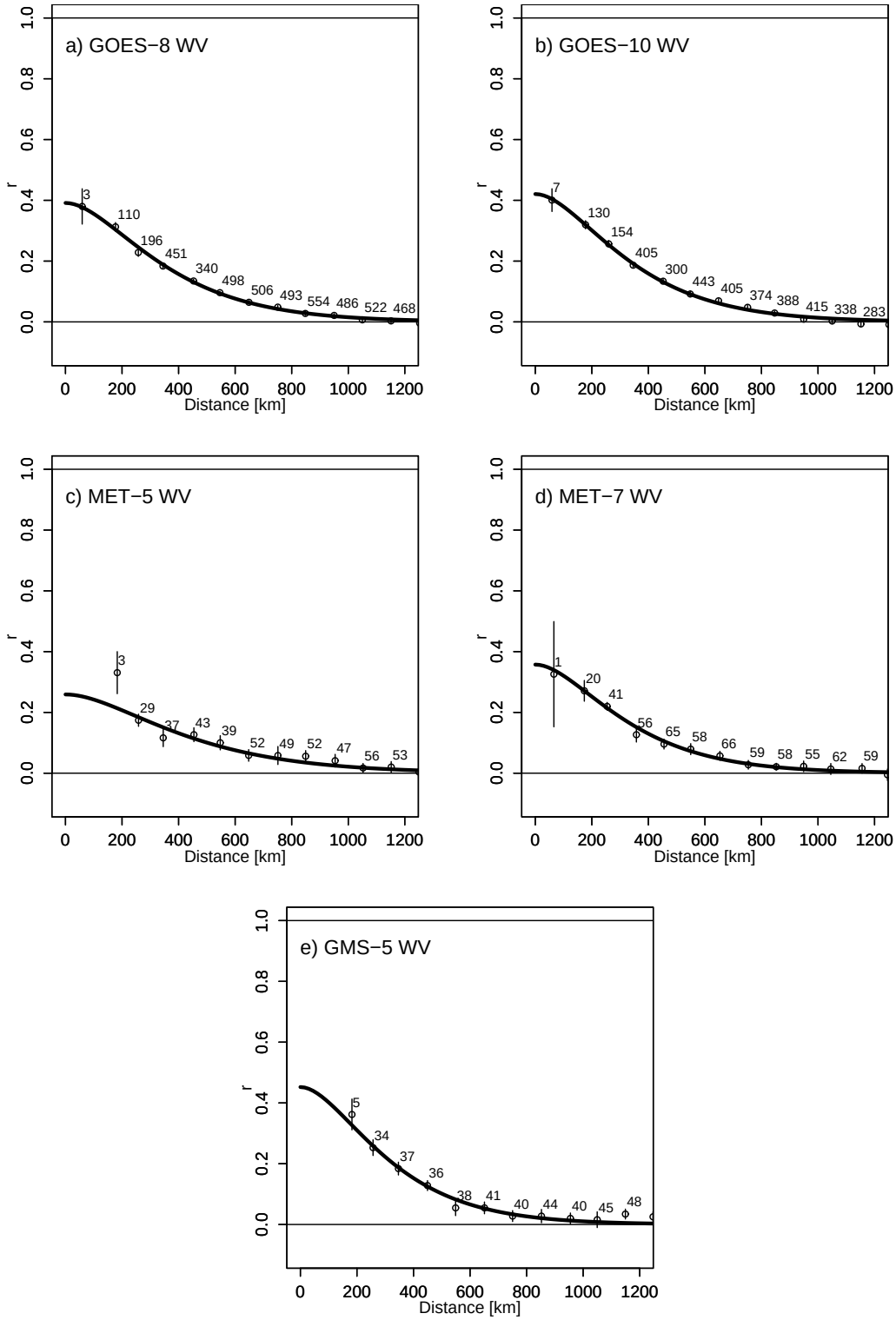

Figure 2: a) $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ AMV-sonde departure correlations for Northern Hemisphere GOES-8 WV winds from all levels as a function of station separation. Error bars indicate 95 % confidence intervals calculated as described in the 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 MET-5 WV cloud track winds. d) As a), but for MET-7 WV cloud track winds. e) As a), but for GMS-5 WV winds.

ie., at scales much larger than the typical size of a processing segment (e.g., 160 km at nadir in the case of EUMETSAT winds, Schmetz et al. 1993). 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 fitted correlation functions fit the correlation data extremely well, mostly within the 95% confidence interval shown, except for the closest distance point for MET-5 data.

The fitting parameters obtained for the correlation function show no statistically significant differences for the five WV AMV datasets, except for Northern Hemisphere MET-5 winds which show a significantly larger L and a smaller intercept R_0 (Table 2). 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. For instance, one might expect larger correlation scales for GOES winds where AMVs are adjusted by comparing them against other observations and forecast data in the "autoeditor" (e.g., Hayden and Velden 1991). While Northern Hemisphere correlation scales for GOES WV winds indeed tend to be larger than for MET-7 or GMS-5 winds, the differences are not statistically significant. However, it should be stressed that 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 results similar to the WV winds with fitting parameters

Table 3: As Table 2, but for IR winds.

Satellite	R_0	L [km]	σ_{AMV} [m/s]
NH: all levels			
GOES-8	$0.40^{+0.02}_{-0.02}$	220^{+8}_{-8}	$3.3^{+0.1}_{-0.1}$
GOES-10	$0.39^{+0.03}_{-0.03}$	207^{+12}_{-12}	$3.1^{+0.1}_{-0.1}$
MET-5	$0.34^{+0.22}_{-0.05}$	190^{+60}_{-50}	$2.7^{+0.8}_{-0.2}$
MET-7	$0.38^{+0.06}_{-0.05}$	185^{+21}_{-20}	$2.8^{+0.3}_{-0.2}$
GMS-5	$0.33^{+0.20}_{-0.07}$	280^{+130}_{-80}	$2.7^{+0.7}_{-0.2}$
NH: above 400 hPa			
GOES-8	$0.40^{+0.03}_{-0.02}$	206^{+9}_{-9}	$3.3^{+0.1}_{-0.1}$
GOES-10	$0.36^{+0.03}_{-0.03}$	201^{+10}_{-8}	$3.0^{+0.1}_{-0.1}$
MET-5	$0.35^{+0.30}_{-0.15}$	150^{+60}_{-30}	$2.7^{+1.0}_{-0.5}$
MET-7	$0.29^{+0.13}_{-0.05}$	170^{+50}_{-30}	$2.7^{+0.6}_{-0.4}$
GMS-5	$0.29^{+0.19}_{-0.07}$	320^{+100}_{-110}	$2.8^{+1.0}_{-0.4}$
NH: 400–700 hPa			
GOES-8	$0.46^{+0.04}_{-0.04}$	200^{+9}_{-9}	$3.3^{+0.2}_{-0.1}$
GOES-10	$0.42^{+0.03}_{-0.02}$	208^{+10}_{-11}	$3.1^{+0.1}_{-0.1}$
MET-7	$0.44^{+0.14}_{-0.09}$	130^{+40}_{-10}	$2.9^{+0.5}_{-0.4}$
NH: below 700 hPa			
GOES-8	$0.42^{+0.12}_{-0.06}$	201^{+27}_{-27}	$2.8^{+0.4}_{-0.2}$
GOES-10	$0.41^{+0.11}_{-0.05}$	210^{+50}_{-30}	$2.7^{+0.4}_{-0.2}$
MET-7	$0.51^{+0.08}_{-0.11}$	210^{+80}_{-20}	$2.9^{+0.3}_{-0.4}$
Tropics: all levels			
GOES-8	$0.27^{+0.11}_{-0.07}$	340^{+100}_{-20}	$2.2^{+0.5}_{-0.3}$
MET-5	$0.29^{+0.20}_{-0.10}$	180^{+190}_{-50}	$2.4^{+0.7}_{-0.5}$

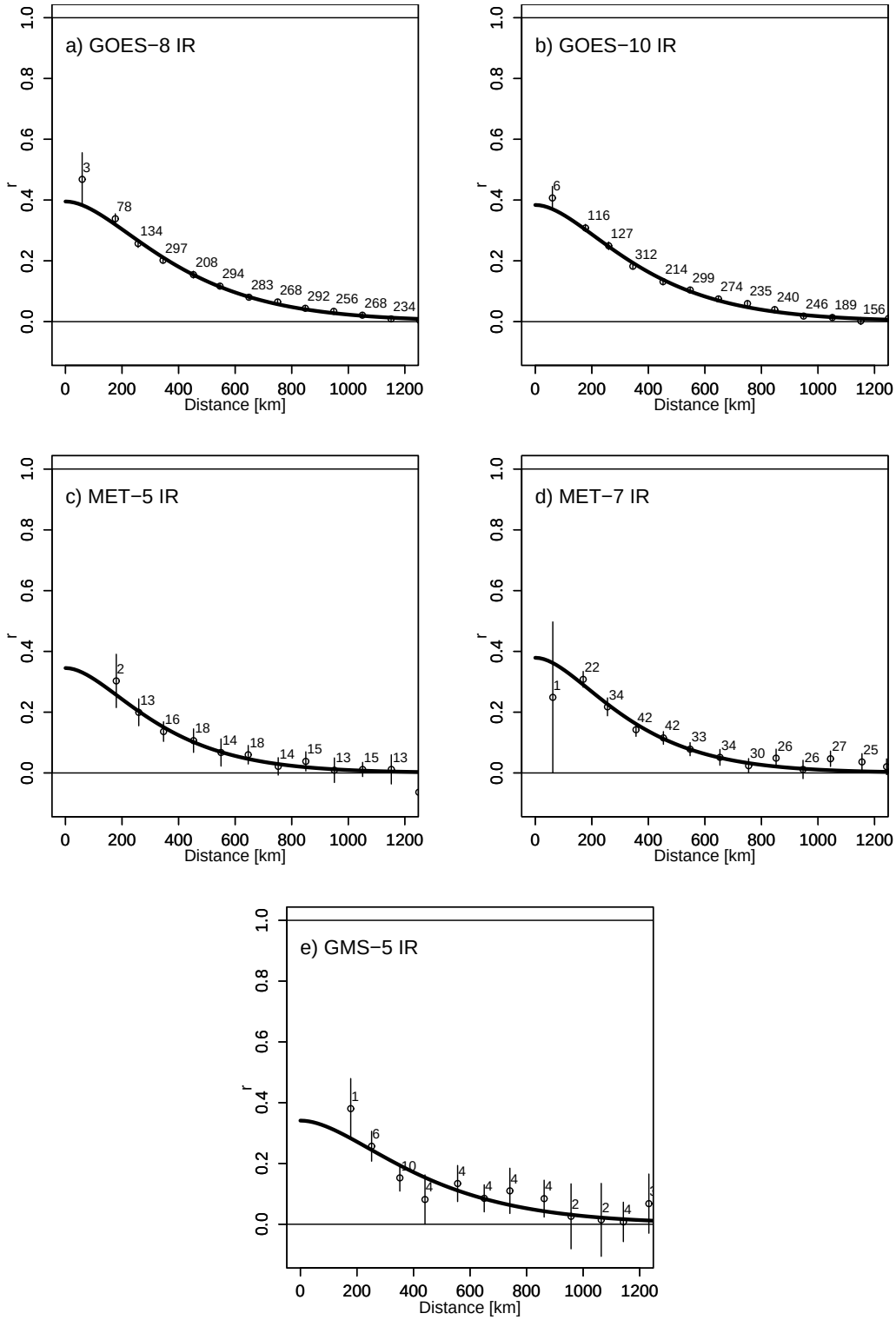

Figure 3: As Fig. 2, but for IR winds.

which are not significantly different from their WV counterparts (Fig. 3, Table 3). Note that much fewer collocations are available for IR winds as these cover a larger vertical range leading to fewer matching collocation pairs. The results are therefore less accurate.

There are some indications of latitudinal differences in the error correlation structures (e.g., Fig. 4a and b, Tables 2 and 3). The correlations tend to be flatter and broader over the Tropics for most winds, and these differences in correlation scales are mostly statistically significant, even though results for the Tropics are less accurate. In the Tropics, a typical value for L is 330 km compared to 190 km for the Northern Hemisphere. 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 (Table 2).

Error correlation structures derived for different vertical layers do not differ significantly, in the sense that the correlation scales of the fitted correlation function are usually not significantly different (e.g., Fig. 4c, d and e, Tables 2 and 3). Nevertheless, the correlations of the AMV-sonde differences tend to be slightly larger for medium or low level winds (400–700 hPa, 700–1100 hPa) compared to high level winds (0–400 hPa) at distances below 500 km, leading to slightly larger intercepts for the fitted correlation functions. This indicates a different partitioning of the correlated and uncorrelated error contributions in the AMV-sonde differences, with slightly stronger spatially correlated contributions at medium and lower levels.

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. It should be pointed out that the correlation function acts as a filter and determines the relative representation of smaller or larger scales in the fit. This in turn somewhat arbitrarily affects the intercept and thus the partitioning of the AMV-sonde variance into correlated and uncorrelated parts. Similar points have been raised by Hollingsworth and Lönnberg (1986). It is reassuring, however, that different correlation functions usually gave results for the AMV errors within the uncertainty intervals shown.

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. The differences between the other four wind datasets are not statistically significant. Also, within the accuracy of the method used, errors for the Southern Hemisphere GMS-5 WV winds show similar values. The errors for the tropical winds tend to be significantly lower, with values of 1.6–2.5 m/s. For reasons discussed earlier, the estimates for the AMV errors come with large uncertainties in regions outside the Northern Hemisphere. GOES WV winds below 400 hPa have significantly larger errors than above 400 hPa, reflecting the difficulty in deriving AMVs from WV imagery in clear-sky and near-no-cloud regions.

IR winds give similar, but slightly smaller errors with overall Northern Hemisphere errors of around 2.7–3.3 m/s. For GOES or MET-7 IR winds there is little difference in the AMV error estimates at different levels. This means that for low-level IR winds, the errors appear surprisingly large - larger than, for instance, typical departures against the First Guess for the winds used in the ECMWF assimilation system. The most likely explanation is that the statistics presented here are derived over land, whereas low level winds in the ECMWF system are used over sea only. Our results for low-level winds are thus not representative for the data typically used in assimilation systems. Also, the representativeness error between the sondes (point measurement) and the AMVs (area average) may not be entirely spatially uncorrelated for lower levels as it is influenced by the lower boundary (Appendix A). This would lead to too large AMV-sonde departure correlations and therefore larger estimates of the correlated error.

An investigation of GOES-8 and GOES-10 winds separately for Northern Hemisphere winter and summer months shows larger spatially correlated errors in winter (Table 4). This partly reflects stronger jet streams

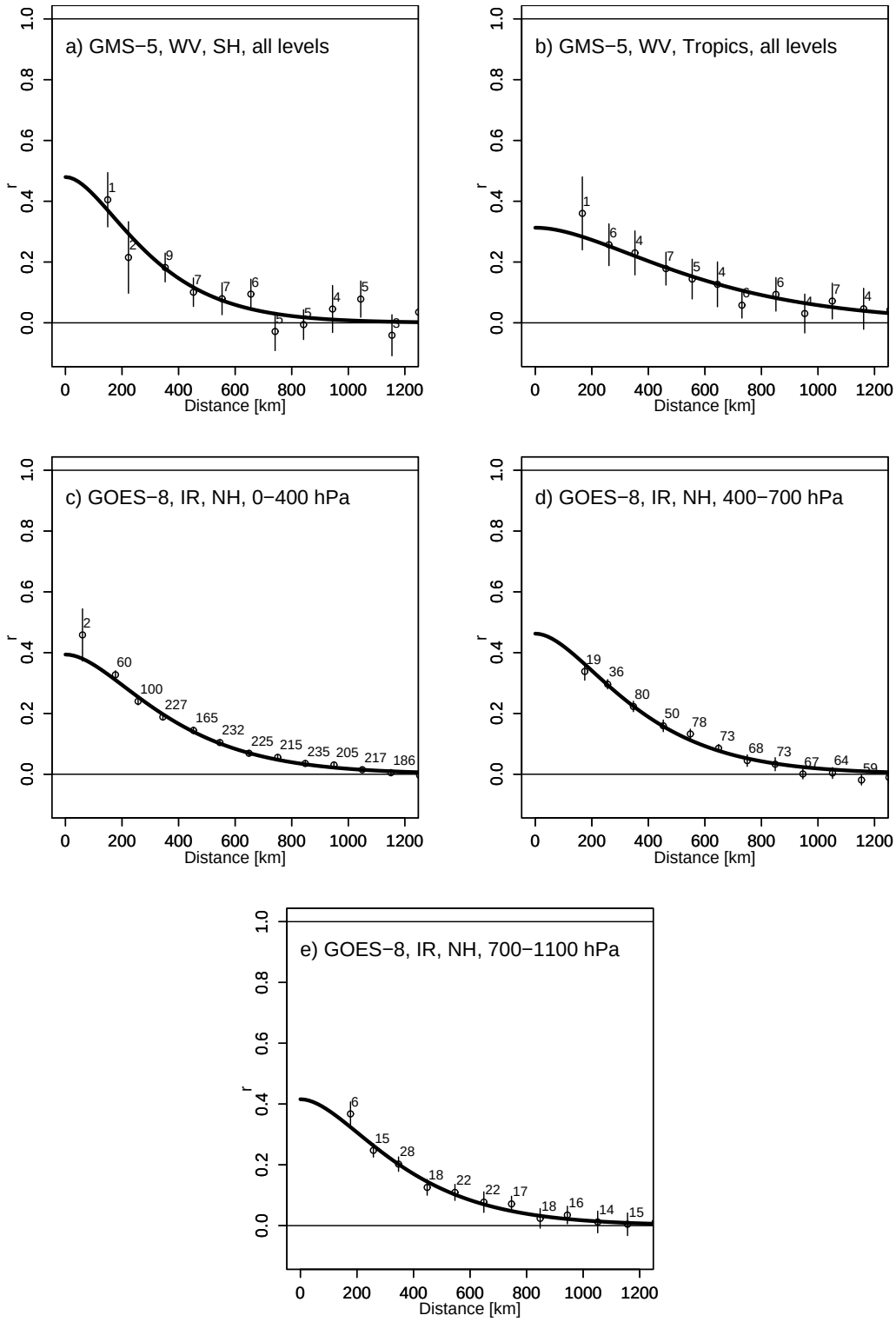

Figure 4: a) $\frac{1}{2}(\langle \Delta u, \Delta u \rangle + \langle \Delta v, \Delta v \rangle)$ AMV-sonde departure correlations for Southern Hemisphere GMS-5 WV winds from all levels as a function of station separation. Error bars indicate 95 % confidence intervals calculated as described in the 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 tropical GMS-5 WV winds. c) As a), but for Northern Hemisphere GOES-8 IR winds above 400 hPa. d) As c), but for winds between 400 and 700 hPa. e) As c), but for winds below 700 hPa.

Table 4: As Table 2, but for GOES-8 and GOES-10 IR and WV winds for Northern Hemisphere winter (December–February) and Northern Hemisphere summer (June–August).

Satellite	R_0	L [km]	σ_{AMV} [m/s]
NH winter, WV, all levels			
GOES-8	$0.47^{+0.05}_{-0.04}$	183^{+13}_{-12}	$4.0^{+0.2}_{-0.2}$
GOES-10	$0.49^{+0.05}_{-0.03}$	193^{+10}_{-10}	$4.0^{+0.2}_{-0.2}$
NH summer, WV, all levels			
GOES-8	$0.27^{+0.04}_{-0.02}$	187^{+12}_{-12}	$2.6^{+0.2}_{-0.1}$
GOES-10	$0.35^{+0.05}_{-0.04}$	160^{+15}_{-12}	$2.9^{+0.2}_{-0.2}$
NH winter, IR, all levels			
GOES-8	$0.45^{+0.03}_{-0.02}$	225^{+11}_{-11}	$3.5^{+0.2}_{-0.2}$
GOES-10	$0.43^{+0.05}_{-0.04}$	195^{+19}_{-11}	$3.3^{+0.2}_{-0.2}$
NH summer, IR, all levels			
GOES-8	$0.28^{+0.03}_{-0.03}$	200^{+30}_{-20}	$2.6^{+0.2}_{-0.2}$
GOES-10	$0.32^{+0.08}_{-0.10}$	169^{+20}_{-20}	$2.6^{+0.2}_{-0.1}$

with higher wind speeds and therefore larger variability in winter compared to summer months. However, it is important to note that the larger correlated errors also arise from a significantly different partitioning of the AMV-sonde differences into correlated and uncorrelated contributions, with larger spatially correlated contributions in winter. This is shown by larger intercepts of the correlation function. Variability is thus not the only reason for the seasonal differences in correlated errors.

The estimates for the correlated part of the high and medium level 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 the errors currently assigned in the ECMWF assimilation system (Table 5), and they are also lower than observation errors used at most other 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 %.

Table 5: Observation errors assigned to AMVs in the ECMWF system.

Pressure level [hPa]	1000–700	500	400	≤ 300
Observation errors [m/s]	2.0	3.5	4.5	5.0

3.3 Anisotropic error correlations

We will now briefly explore the anisotropic aspects of the error correlations. This exercise is particularly dependent on the sample size as it requires 2-dimensional grouping of the data in W–E and S–N direction. Therefore, meaningful statistics could only be produced for Northern Hemisphere WV winds, and Northern Hemisphere IR winds from GOES-8 and GOES-10. Also, the binning size required for meaningful statistics (300-400 km boxes) means that the correlations appear considerably smoothed.

The statistics 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

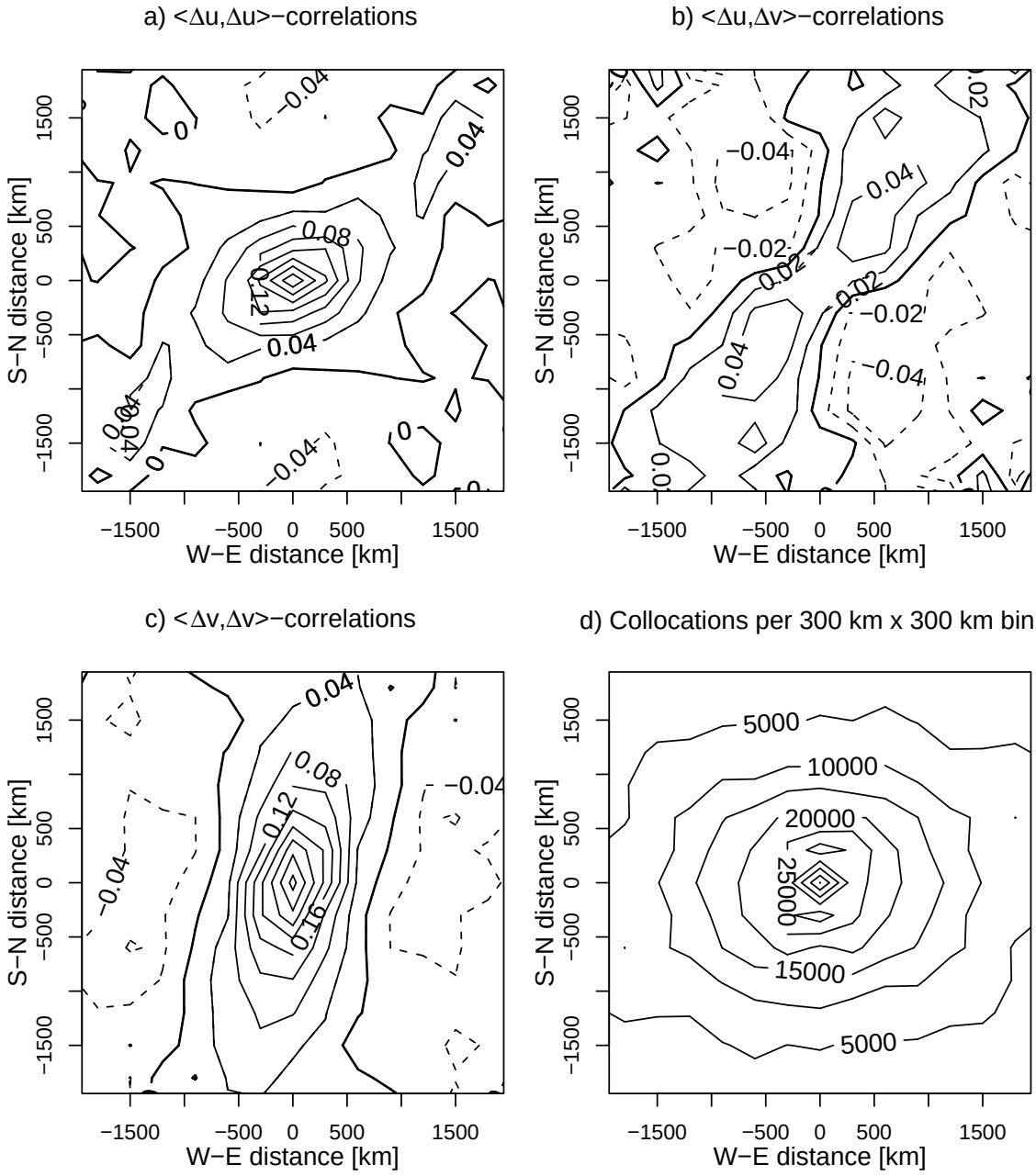

Figure 5: a) Spatial structure of the $\langle \Delta u, \Delta u \rangle$ AMV-sonde departure correlations for Northern Hemisphere GOES-8 WV winds as a function of W–E and S–N distance from an AMV located at (0 km,0 km). Note that only the upper half of the plot was computed from the available data; the lower half was obtained through point symmetry. The binning box size is 300 km $\times$ 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$\times$300 km bin.

WV or IR winds considered (e.g., Fig. 5), except for GMS-5 WV winds (not shown). Apart from a slightly more diagonal orientation, the correlation structures shown in Fig. 5 agree well with theoretical correlations for non-divergent flow in the case of isotropic correlations for the velocity potential and streamfunction (e.g., Daley 1993). The anisotropic structure thus suggests that non-divergent errors dominate in extra-tropical AMVs. Such behaviour is also typical for short-term forecast errors (e.g., Daley 1993, Hollingsworth and Lönnberg 1986).

3.4 Error covariance of an idealised AMV dataset

To highlight some of the implications of the above findings on the error covariance matrix for AMVs we will now investigate the error correlation matrix of a hypothetical idealised dataset. For simplicity we investigate a scalar quantity. Consider observations on a regular grid covering a flat 10,000 km $\times$ 10,000 km domain. The domain size for this exercise was chosen to be comparable to that covered by AMVs from one geostationary satellite, but the results were found to be mainly insensitive to this size as long as it is much larger than the correlation scales involved. The correlation matrix $\mathbf{O}$ of this dataset is calculated assuming isotropic correlations given by the correlation function R with a correlation length scale of L=190 km, typical for high-level midlatitude AMVs. Assuming uniform observation errors, an eigen-decomposition of the error correlation matrix $\mathbf{O}$ allows us to study the spatial structures of the observation error of the whole dataset. The eigenvalues of $\mathbf{O}$ characterise the error variance (normalised by the uniform observation error variance) associated with the corresponding spatial structures given by the orthogonal eigenvectors. Doing this decomposition for different grid spacings of the hypothetical dataset characterises the impact of the AMV resolution (as controlled, for instance, by thinning in a data assimilation system) on $\mathbf{O}$ compared to a diagonal matrix.

Table 6 shows how the range of eigenvalues of $\mathbf{O}$ expands as the AMV spacing approaches the correlation length scale (For ease of interpretation we give the square root of the eigenvalues, corresponding to the normalised errors themselves rather than the error variances.). In contrast, if the errors were uncorrelated all these spatial structures would have the same observation error and eigenvalues of 1. The error of mean-like structures associated with the leading eigenvector (e.g., Fig. 6) is significantly increased with correlated errors compared to the case of a diagonal covariance matrix. In other words, if error correlations are taken into account, such structures in the observations will receive less weight in an assimilation, compared to an assimilation in which such correlations are neglected. In contrast, smaller scale structures have smaller errors in the case of correlated errors, and an analysis system which uses these correlated errors will put more weight on these observational structures than one which assumes uncorrelated errors. Similar findings have been reported by Seaman (1977) who showed that correlated observation errors increased the accuracy of the analysis of gradients while reducing the accuracy of the absolute quantities, compared to uncorrelated observation errors. Our analysis also highlights that such differences in the errors associated with the eigen-structures can not be accounted for by simply inflating the observation error and using a diagonal covariance matrix.

It should be stressed here that it is difficult to judge the relevance of the above findings for actual data assimilation systems, and that it is beyond the scope of this study to assess this issue. While the above considerations provide some insight into the nature of the observation error covariance matrix for an AMV dataset, the filtering properties of the analysis system will be further influenced by, for instance, background errors and the analysis resolution. Also, the regular AMV coverage and the error correlations are highly idealised, whereas in a realistic dataset the coverage will depend on the available tracers in the satellite imagery and the correlations have latitudinal and anisotropic variations.

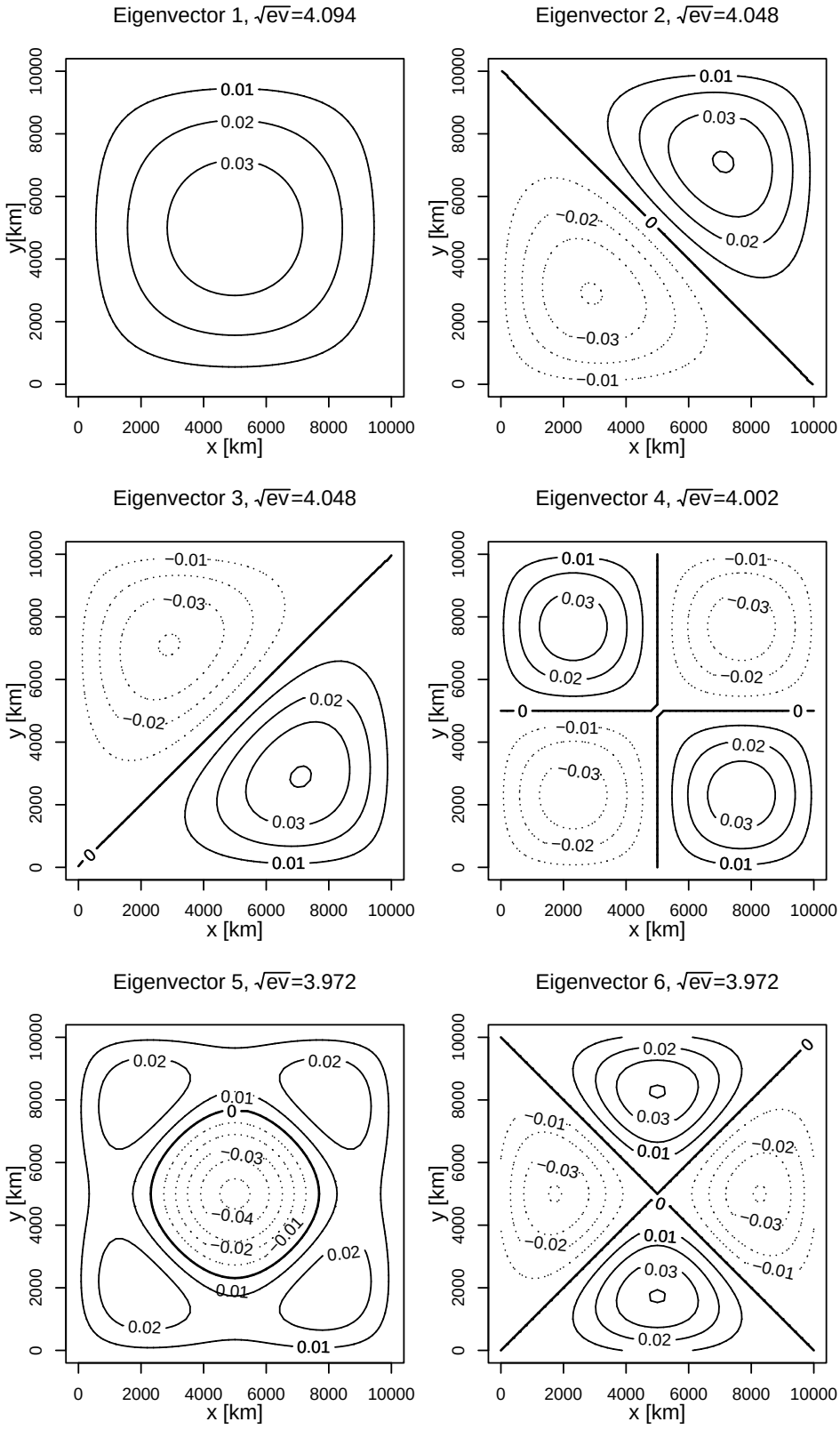

Figure 6: The leading six eigenvectors of the error correlation matrix of an idealised observational dataset covering a 10,000 km$\times$10,000 km domain with a regular 200 km spacing. The error correlations are given by the correlation function R with a length scale of $L = 190$ km. The square root of the corresponding eigenvalue is also shown, corresponding to the error inflation factor for this eigen-structure. See main text for further details.

Table 6: Range of the square root of the eigenvalues of **O** and number of available observations for an idealised observational dataset at different regular spacings Δx, covering a 10,000 km$\times$10,000 km domain. The value of the correlation function R at the grid spacing Δx is also shown. See main text for more details.

Δx [km]	$R(\Delta x)$	Number of AMVs	Square root of the largest and smallest few eigenvalues of **O**
800	0.077	169	1.17, 1.17, ..., 0.88, 0.88
600	0.177	289	1.43, 1.42, ..., 0.74, 0.74
400	0.378	676	2.06, 2.04, ..., 0.51, 0.51
200	0.716	2601	4.09, 4.05, ..., 0.22, 0.22
150	0.813	4489	5.46, 5.39, ..., 0.14, 0.14

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 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. There is evidence for larger correlated errors in the extra-tropics in winter compared to summer. 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 structures are consistent with correlations in cases where errors in the non-divergent wind dominate.

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), and temperature forecast errors also tend to exhibit broader correlations in the tropics (e.g., Daley 1993). Similarly, the anisotropic structures observed in the AMV error correlations in the extra-tropics are typical of short-term wind forecasts (e.g., Daley 1993). As errors in the forecast data are also inherently temporally correlated, AMVs are likely to exhibit temporal as well as spatial error correlations.

The findings about the spatial correlations of the AMV errors highlight some shortcomings in the use of AMVs at many data assimilation centres. As previously suspected, 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). This explains the need to use inflated observation errors at most NWP centres (e.g., Tsuyuki 2000), as the information content of the data is lower than if all AMVs were independent (e.g., Bergman and Bonner 1976). However, it should be pointed out here that 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 (e.g., Liu and Rabier 2002) and by practical experience with positive forecast impact from the assimilation of AMVs with such suboptimal settings (e.g., Bouttier and Kelly 2001, Rohn et al. 2001).

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. It should be stressed that without a careful analysis of the filtering properties of an analysis system it is impossible to assess to what extent error correlations in AMVs affect their assimilation. It is thus beyond the scope of this study to recommend which of the above options is the most cost-effective approach. Nevertheless, given the significant correlations, it appears beneficial to explicitly account for spatial (and possibly temporal) error correlations for AMVs in data assimilation. This is likely to be particularly true for mesoscale assimilation. The use of correlated AMV errors is technically not straightforward, as it requires, for instance, the inversion of very large non-diagonal observation error covariance matrices. If such an approach is therefore technically not possible this study provides guidance for a careful revision of the combination of thinning and inflated observation errors. Our results highlight how thinning and error inflation are a compromise between excluding information, but satisfying assumptions on the error characteristics on the one hand, and using more data, but misrepresenting spatial error characteristics on the other. It should also be pointed out that AMVs show considerable speed biases (e.g., Bormann et al. 2001), and our error characterisations use AMVs after a bias removal. The speed biases may demand other adjustments to the observation errors.

The significant error correlations for separations of up to about 800 km also have some implications on the resolution at which AMVs should be derived by winds producers. Today's wind datasets provide winds at 160 km resolution or well below this. Such high-resolution data allows a detailed representation of flow structures and remains useful in nowcasting applications or to derive other quantities such as divergence from the AMV field. Nevertheless, such applications would benefit from a better characterisation of the actual information content of the AMV data. For data assimilation applications, improvements in other aspects of the AMV product, for instance for the common speed bias in the AMVs, are likely to be more beneficial than further increases in the resolution. The strong correlations also highlight the importance of providing quality information with the derived high-resolution product, so that the "best" wind within a region can be selected in the assimilation if correlated observation errors cannot be explicitly accounted for (e.g., Holmlund et al. 2001, Rohn et al. 2001).

Appendix A

Assuming zero biases, the correlation between the AMV-sonde departure pairs for each wind component is given by

$$\rho = \frac{\langle (a_1 - o_1)(a_2 - o_2) \rangle}{\sqrt{\langle (a_1 - o_1)^2 \rangle \langle (a_2 - o_2)^2 \rangle}} \tag{3}$$

with AMV values a_1 and a_2 and sonde observations o_1 and o_2 from location 1 and 2, respectively. The angle brackets indicate summation over all collocation pairs for these two sites. We now introduce the true values $\hat{a}_1$, $\hat{a}_2$, $\hat{o}_1$, and $\hat{o}_2$ for the observations, respectively. Note that the "truth" may differ for AMVs and sondes given the different representativeness of the two observations. The numerator becomes

$$\begin{aligned}
&\langle (a_1 - o_1)(a_2 - o_2) \rangle \\
&= \langle (a_1 - \hat{a}_1 + \hat{a}_1 - o_1 + \hat{o}_1 - \hat{o}_1)(a_2 - \hat{a}_2 + \hat{a}_2 - o_2 + \hat{o}_2 - \hat{o}_2) \rangle \\
&= \langle (a_1 - \hat{a}_1)(a_2 - \hat{a}_2) \rangle + \langle (o_1 - \hat{o}_1)(o_2 - \hat{o}_2) \rangle + \langle (\hat{a}_1 - \hat{o}_1)(\hat{a}_2 - \hat{o}_2) \rangle \\
&\quad + 2 \langle (a_1 - \hat{a}_1)(\hat{a}_2 - \hat{o}_2) \rangle - 2 \langle (o_1 - \hat{o}_1)(\hat{a}_2 - \hat{o}_2) \rangle \\
&\quad - 2 \langle (a_1 - \hat{a}_1)(o_2 - \hat{o}_2) \rangle
\end{aligned} \tag{4}$$

where it has been assumed that mixed correlations such as between the AMV error $(a_1 - \hat{a}_1)$ and the error of representativeness $(\hat{a}_2 - \hat{o}_2)$ are the same for location 1 and 2.

The first term on the right hand side depends on the spatial error correlations of the AMVs, and it is assumed to be the main contributor to non-zero correlations for station distances greater than zero. The second term relates to the spatial error correlations of the sondes and it is assumed to be zero as long as the two sondes differ, following the assumption on spatially uncorrelated sonde errors. The third term depends on the spatial correlation of the errors of representativeness. We assume that this correlation is small for station distances greater than zero, but for lower level flow affected by the lower boundary it may be non-negligible. The fourth and fifth terms depend on the correlations between the AMV or sonde error and the error of representativeness (including collocation errors), respectively. It is reasonable to assume that both are zero. The last term is determined by the correlations between the AMV error and the sonde error. For EUMETSAT and JMA data it is reasonable to assume that the two are uncorrelated as the sondes are not used in the processing. For NOAA/NESDIS data other observations enter the processing during the quality control, so correlations are possible. However, these are expected to be small.

Acknowledgements

The study benefited from constructive comments and suggestions from Erik Andersson, Mike Fisher, Anthony Hollingsworth, and Adrian Simmons.

References

Bergman, K. H., and W. D. Bonner, 1976: Analysis error as a function of observation density for satellite temperature soundings with spatially correlated errors. *Mon. Wea. Rev.*, **104**, 1308–1316.

Bormann, N., G. Kelly, and J.-N. Thépaut, 2001: Characterising and correcting speed biases in atmospheric motion vectors within the ECMWF system. In *Proc. 2001 EUMETSAT Meteorological Satellite Data User's Conf.*, Antalya, Turkey, EUMETSAT, 596–603.

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.

Hayden, C. M., and C. S. Velden, 1991: Quality control and assimilation experiments with satellite derived wind estimates. In *Proc. Ninth Conf. Numerical Weather Prediction*, Denver, CO, USA, Amer. Meteor. Soc., 19–23.

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.

Holmlund, K., C. Velden, and M. Rohn, 2001: Enhanced auomated quality control applied to high-density satellite-derived winds. *Mon. Wea. Rev.*, **129**, 517–529.

Järvinen, H., and P. Undén, 1997: Observation screening and background quality control in the ECMWF 3DVAR data assimilation system. Technical Memorandum 236, ECMWF, Reading, U.K.

Liu, Z.-Q., and F. Rabier, 2002: The interaction between model resolution and observation resolution and density in data assimilation: A one-dimensional study. *Quart. J. Roy. Meteor. Soc.*, **128**, accepted.

Lorenc, A. C., S. P. Ballard, R. S. Bell, N. B. Ingleby, P. L. F. Andrews, D. M. Barker, J. R. Bray, A. M. Clayton, T. Dalby, D. Li, T. J. Payne, and F. W. Saunders, 2000: The Met. Office global 3-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **125**, 2991–3012.

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.

Seaman, R., 1977: Absolute and differential accuracy of analyses achievable with specified observation network characteristics. *Mon. Wea. Rev.*, **105**, 1211–1222.

Soden, B. J., C. S. Velden, and R. E. Tuleya, 2001: The impact of satellite winds on experimental GFDL hurricane model forecasts. *Mon. Wea. Rev.*, **129**, 835–852.

Thiébaux, H. J., 1976: Anisotropic correlation functions for objective analysis. *Mon. Wea. Rev.*, **104**, 994–1002.

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/sec5/ nwp/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.

Weber, R. O., and P. Talkner, 1993: Some remarks on spatial correlation function models. *Mon. Wea. Rev.*, **121**, 2611–2617.