ADM-AEOLUS MAG NOTE

DATE: 21-6-2002; REVISED 6-11-2006

TO: MAG DISTRIBUTION LIST

FROM: AD STOFFELEN; GERT-JAN MARSEILLE

SUBJECT: SPATIAL REPRESENTATIVENESS

BACKGROUND

Spatial representativeness is an important concept for ADM since it determines to a substantial extent the performance. The concept is discussed extensively by Lorenc et al (1992), and later on it is used by Stoffelen et al (1994) in the ADM simulator. Stoffelen (1998) discusses the requirements for ADM in the context of the meteorological data assimilation problem and in particular explains the spatial scales involved. Based on all this, lidar burst mode operation was brought forward. This note intends to clarify and summarise the issue of spatial representation.

Spatial representation is important for the understanding of observation error contributions in data assimilation, but also to amend the spatial scales of the nature run in an OSSE, in order to simulate the observations from the nature run realistically. Both issues are elaborated on below.

NATURE RUN

In preparation for an OSSE, it is important to add realistic small scales when simulating "true" highly-resolving observations in order to guarantee the OSSE realism. True observations are obtained by interpolating the nature run fields towards the observation location and transforming "nature" run state variables to observed variables. The smallest possible spatial scales represented by the nature run are determined by the truncation of the model that generated the nature run, i.e. T511/T799 (40/25 km) for the NCEP-OSSE. Observations in the real world may include still smaller scales and these should be accounted for in the simulation of the observations by adding a representativeness (random) process with known variability. The magnitude of this variability contribution accounts for the energy represented by scales smaller than those represented in the nature run, but still resolved by the observing system (radiometrically, spatially, temporally, etc.). A way to quantify this error is described in the next section. A second contribution lies in the geophysical transformation from "nature" run state variables to observed variables. This operation may be scale (resolution) dependent.

In summary, representativeness variability contributions to be added to simulated true observations are directly related to the numerical characteristics of the "nature" run model, e.g., grid cell size, horizontal diffusion scheme, cloud parameterisation, etc, and to the scale-dependency of the observation operator.

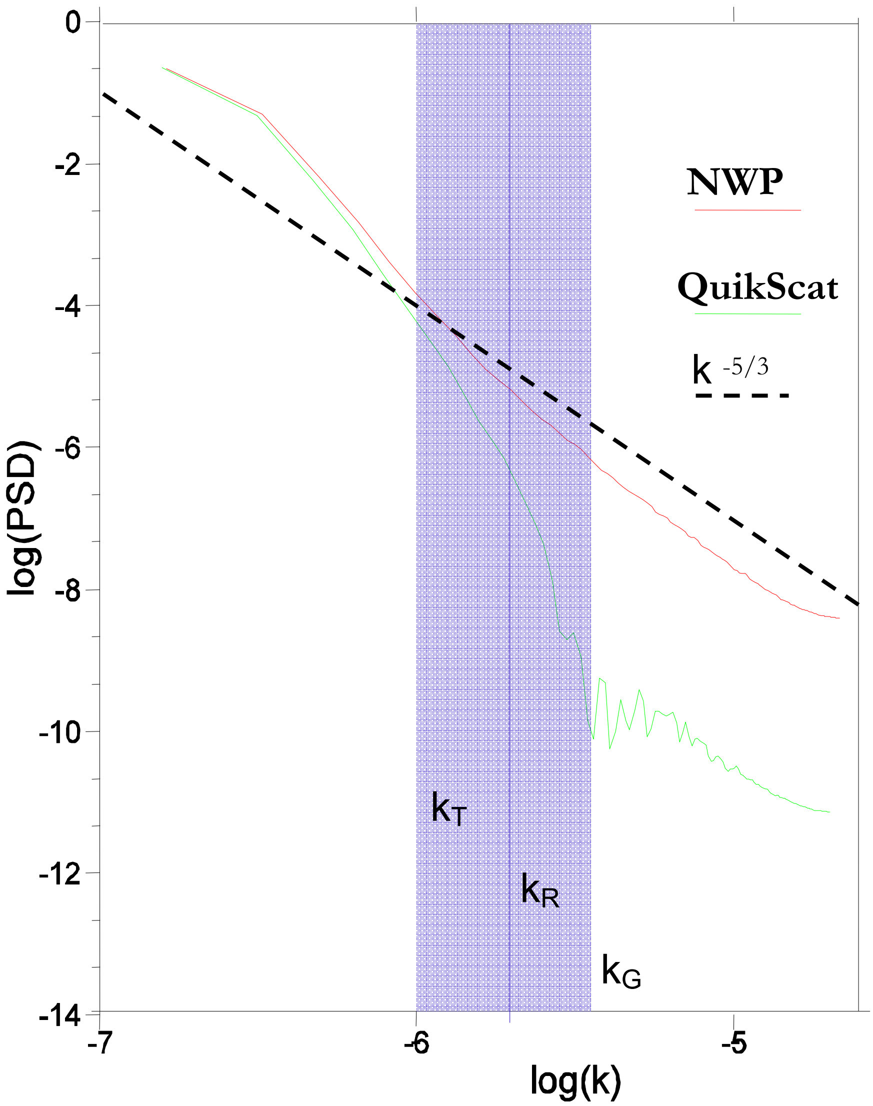

FIGURE: ILLUSTRATION OF REPRESENTATIVENESS ERROR MODEL. THE CURVES RERPRESENT A SPECTRAL SHAPE OF $K^{-5/3}$, AN ESTIMATE OF THE KNMI SEAWINDS 25-KM WIND PRODUCT, AND THE COLLOCATED NCEP 1000-MB WIND SPECTRAL CONTENT IN LOG-LOG SCALE (VOGELZANG, 2006). WITH RESPECT TO THE TWO OTHER CURVES, THE NWP SPECTRAL CURVE DROPS SEVERAL ORDERS OF MAGNITUDE IN VARIANCE OVER THE BLUE SHADED SPECTRAL RANGE, CHARACTERISED BY THE SPECTRAL WAVE NUMBERS INDICATED (SEE TEXT) RANGING FROM CORRESPONDING WAVELENGTHS OF 1000 KM TO .250 KM.

DATA ASSIMILATION

In meteorological analysis, the atmospheric state is represented in grid boxes. However, NWP models do generally not contain the smallest possible scales and are spatially and temporally truncated. Let's therefore leave the concept of a grid box and presume that we perform analysis on the "resolved" NWP model scales, called here "resolution cell". In each resolution cell, the best mean meteorological state is sought for. Observations, like those from ADM, do generally not provide volume-mean quantities. As such spatial representation errors exist. For a point measurement, like from a radiosonde, the spatial representativeness error in a wind component is estimated to be 2-3 m/s, depending on height (Stoffelen et al, 1994). If the resolution cell size decreases, the spatial representativeness error of the point measurement should decrease, since less wind variability is captured by the cell. This can be verified by looking at differences between the NWP model state and such point observations (e.g., o-b statistics). Note however that both the background and the observation representativeness errors change when NWP model variability (resolution) changes due to the error model defined below. Generally, nowadays a horizontal resolution cell of ~150 km may be taken as representative of global NWP models.

SPECTRAL TRUNCATION

Lorenc et al. (1992) describes wind component variability by a wind component energy density spectrum following both experimental and theoretical evidence

$$E(k)dk = E_0 k^{-5/3} dk \qquad (1)$$

where k is wave number. Integration from wavenumber k_{RC} , the wave number of the resolution cell, to infinity (0 wavelength) gives the horizontal component wind variability on the scales not resolved by the NWP model as

$$r^2 = \frac{3}{2} E_0 k_{RC}^{-2/3} \qquad (2)$$

To estimate the NWP model truncation, a model spectrum may be compared to this spectrum (Eq.1) at the 1000-2000 km scale to obtain E_0 and the missing variance substituted in Eq. (2) to obtain k_{RC}. The spectral gap of the NWP model centered around k_{RC} will in general extend from k_T, the lowest wavenumber where truncation is present, to k_G, the highest wave number with hardly any variance, usually corresponding to a few times the grid box size (see figure). We further note that as grid lengths are decreased, natural forcing will in some areas (e.g., with orography) provide additional deterministic mesoscale structures, but not in areas with weak forcing (e.g., over the oceans as in the figure). Here, 3D observations will be needed to provide an improved deterministic mesoscale flow.

ADM-AEOLUS REPRESENTATIVENESS ERROR

The ADM DWL is side-looking and describes a linear track on the Earth's surface. We distinguish the along-track and across-track wind component variability, where the wind component is taken in the direction of the LOS. We assume that the along- and across-track variabilities are independent, but of equal size in amplitude, i.e., r^2.

In the across-track direction we only obtain one measurement location, and as such the wind variability in this direction is not resolved. The LOS wind component representativeness error thus amounts to r^2.

In the along-track direction, over a length of 50 km, we may obtain up to N (e.g., 14) accumulations from ADM. 50 km represents a third of the resolution cell, which for a $k^{-5/3}$ variability spectrum, is about half of the variability. So, if the 14 samples are uniformly spaced, then we effectively sample half of the LOS wind component variability, and reduce its error variance contribution to

$$\left[1-\left(\frac{\text{sample length}}{\text{resolution cell}}\right)^{2/3}\right]r^2+\left(\frac{\text{sample length}}{\text{resolution cell}}\right)^{2/3}\frac{1}{N}r^2$$
$$\approx\frac{1}{2}r^2+\frac{1}{2N}r^2 \qquad (3)$$

for a 50 km along track sample length.

In addition to these inevitable spatial representativeness errors, we unfortunately also have to deal with measurement errors, i.e., m^2. For ADM, the measurement error variance is proportional to the photon count on the detector. m^2 depends on the product of Pulse Repetition Frequency, PRF, and laser shot energy and this product should be optimised. Changing either energy or PRF of the laser by the same relative amount thus has a very similar effect. For N measurements in one observation, the error is reduced to m^2/N.

So, let us define the total error variance as

$$o^2 = 3\,r^2/2 + r^2/2N + m^2/N \qquad (4)$$

Let us take the ADM case with N = 14 and r^2 = 4 m²s⁻² (mid-troposphere case). The requirement for the total wind component observation error variance is based on 2 times r^2 (to simulate the quality of a radiosonde wind). Following eq. 4, this means that $m^2/14$ should be less than $(13/2).r^2/14$, or $m^2 < (13/2).r^2 = 26$ m^2s^{-2} . On the linear observation scale $m^2/14$ equals a bit less than 2 m^2s^{-2} and the linear observation scale observation error thus about 1.4 m/s. In other words, by sampling the along-track direction we reduce the total error variance due to spatial representation by about a 4th, and for meeting a user requirement to match radiosonde quality it may be replaced by measurement error.

As mentioned before, E_0 tends to be somewhat height dependent and maximum at jet level. The ADM performance issue is most critical at a height of 10 km (e.g., ESA 1999).

RESOLUTION CELL AND ACCUMULATION DISTANCE

Ad, behalve het feit dat je analyse hieronder niet kan volgen, vind ik het jammer dat je de discussie niet voert vanuit informatie inhoud: Als ADM sampled over 100 km (2 aangrenzende grid boxen) dan is de bijbehorende fout in de ADM waarneming volgens de bovenstaande formule $o^2 = r^2 + r^2/4 + 3r^2/4N + m^2/N = 7.2$ m^2s^{-2} tegen 8.14 m^2s^{-2} voor 50 km (en voor N=14, m/14=2) ADM sampling oftewel een ruim 10% kleinere fout in de 100 km ADM waarneming. Echter, deze waarneming moet in de analyse opboksen tegen een background estimate die in de variantie bijna 2 keer kleiner is (gemiddelde van de background estimates in de 2 grid boxes). Deze analyse is iets te simpel want de background fouten in de 2 naburige grid boxen zijn gecorreleerd, dus zal de factor 2 wat kleiner zijn, maar toch wel groter dan 1.1 denk ik. Het relatieve gewicht (informatie inhoud) van een 100 km ADM meting in de analyse is dan kleiner dan die voor een 50 km sample.

The 50-km accumulation distance of ADM has been specified as a number that represents λ_G.= 1 / k_G., a cell size of global Numerical Weather Prediction models within which generally little variability exists. For larger scales up to a few 100 km, the resolved spectral density increases in order to saturate to the spectral shape of eq. 1 at k_T., as explained at the end of the section on spectral truncation. So, if ADM accumulated over longer distances then 50 km, then one accumulated observation would represent some averaging of NWP model (and real) variability and depend on more than one 50-km cell.

Suppose, we accumulate $M = N$ samples over 100 km, then the ADM observation will be represented by the average of two grid boxes and the reprentativeness problem across will not be very different. However, along the track the sampled distance in the NWP model state has increased beyond λ_G. by 50 km and thus the spatial representativeness problem due to partly sampled real spatial scales is also extended by 50 km. So, while the along-track representativeness error length increases to 200 km, 100 km of it is sampled. Following Eqs. 1-3, r^2 then increases by $(200/150)^{2/3}$ = 1,21 of which a fraction $(100/200)^{2/3}$ = 0.63 is sampled by 14 samples resulting in a representativeness contribution of 1,21{0.37 + 0.63/14}.r^2 = 0,50.r^2 , which added to the contribution m^2 = 26/14 m^2s^{-2} provides an observation error variance of the accumulated 100-km wind of 3.87 m^2s^{-2} . This would be about the same as the error of the 50-km accumulation with 14 samples!

Since we cannot resolve the 50-km size weather anymore by accumulating over 100 km, this option clearly seems not favourable. Alternatively, combining two neighbour 50-km observations is also difficult, since their representativeness errors are spatially correlated, because the corresponding resolution cells overlap. This observation error correlation would then have to be explicitly formulated in the observation operator. It thus appears better to stick to the cell size accumulation length of 50 km in case of uniform conditions and a λ_G. of 50 km.

CLOUDS AND AEROSOL

In obtaining equation 3, we assumed uniform linear sampling in an accumulation cell. However, due to cloud, we will not often have true uniform sampling in the horizontal. In this case, the along-track representativeness error will increase as samples are missing, and even more so if subsequent samples are missing, due to the spectral characteristics of wind variability. In other words, the cloud pattern will determine the true LOS wind component representativeness error. Moreover, equation 3 assumes a constant m^2, which in reality will be modulated by the cloud and aerosol presence, both aloft and in the range gate under consideration. Due to the varying photon count, the reduction in along-track representativeness error will be **less** effective than presented in equation 3. Moreover, variable aerosol and cloud conditions may in practise be generally associated with cases of increased wind variability.

In the simulations of ADM-Aeolus at KNMI, we have assumed that the cloud cover varies from one measurement (1-3.5 km) to the next, but rather not within measurements. This is motivated by the fact that cloud cover occurrence may be described by a spectral power law similar to equation 2 (Stoffelen et al, 1998). This implies that cloud cover variations from shot to shot within a measurement are generally assumed rather minor as compared to cloud cover variations from measurement to measurement within a stretch of 50 km.

QUALITY CONTROL

As said before, variable aerosol and cloud conditions may in practise be generally associated with cases of increased wind variability. ADM is safeguarded against detrimental impact of unrepresentative data by oversampling the horizontal; each 50-km observation is split into several 1-km scale measurements. As such, the signal (and Doppler shift) fluctuations within the 50-km stretch can in principle be exploited to detect large errors due to spatial representation. It is common practise and common sense to reject these observations, since interpretation of the observation as a mean resolution cell weather quantity is extremely difficult in these variable cases. Data assimilation systems are known to be very sensitive to rejection citeria, i.e., quality control. This means that data assimilation systems will also be sensitive to our ability to detect signal variability within an ADM observation.

NEIGHBOURS

Another matter may be whether a continuous mode would be able to provide good measurements in cases where the burst mode would fail because of clouds. The burst mode is optimised to provide independent good-quality observations to the data assimilation system. Independency is warranted by observing every 200 km, such that neighbouring observations do not correct the same resolution cell of the NWP model. Sticking to the 50-km accumulations, we would in principle obtain four observations within a 200-km stretch. These observations are, however, no longer providing independent information to the NWP model analysis, since the observation increments are spread over many grid boxes in the analysis procedure. Moreover, since the resolution cells corresponding to the observations overlap, their total observation errors computed according to eq.3 would be correlated. As a consequence, in data assimilation systems, it has up to now generally proven difficult to exploit such "dependent" high-resolution observations.

However, if the product of PRF and laser shot energy was the same in burst and continuous mode, one would most likely be able to achieve a favourable effect in case that the nominal burst mode observation would be of poor quality, by for example checking the better quality of the two neighbouring 50-km observations. On the other hand, if the photon count will be decreased in case of continuous mode, then the quality of **all** 50-km observations made by ADM will be decreased, and we verified that this procedure of picking favourable "neighbours" in case of cloudy scenes (only) would not compensate for the general loss of data quality. At KNMI we verified this by using LIPAS in continuous mode with LITE data.

SUMMARY

After providing a model for the total observation error and considering the meteorological data assimilation problem, it is clear that a 50-km accumulation length should be used for both continuous and burst mode scenarios.

The specification for quality on a 50-km observation should not be relaxed. On the one hand, more information could be provided in continuous mode, but up to this date it turns out to be difficult to exploit high resolution data. In fact, data thinning procedures that reject observations that are closeby (within 100 or 200 km) are commonly used nowadays. Although one may anticipate that this situation will improve somewhat in the years to come, it appears unwise to count on it (for ADM).

Quality control capability is important for achieving beneficial impact in NWP analysis. The link between quality control effectiveness and data quality has to be further elaborated in order to make an informed decision on change of ADM specification.

Although, there is as of yet mixed evidence that ADM performance in cloudy scenes is actually most relevant, cloudy scenes may be relatively more important for demonstrating the beneficial impact of ADM than only fully clear scenes. A closer analysis of reduced cloud obstruction effects in continuous mode versus improved quality in burst mode provides evidence in favour of good quality scenes in burst mode (LITE4ADM simulation).

REFERENCES

ESA, 1999, ADM assessment report, Granada.

Lorenc, A. C., R. J. Graham, I. Dharssi, B. Mcpherson, N. B. Ingleby and R.W. Lannon, ″Study of preparation for the use of Doppler Wind Lidar information in meteorological assimilation systems″, Final report on ESA study contract 9063/90.HGE-I, published by U. K. Meteorol. Office, Bracknell, England, 1992.

Stoffelen et al, 1994, Theoretical Studies of the Impact of Doppler Wind Lidar Data - Preparation of a data base, ESA-CR(P)-3943.

Stoffelen, Ad, 1998, ADM user perspective, KNMI contribution to ADM_phaseA, CD present at ESA

Veldman, S.M, H.A. Knobbout, A. Stoffelen, G.J. Marseille and E.A. Kuijpers, 1999, LIPAS, executive summary, ESA, 3-9132/97.

Vogelzang, 2006, On the quality of high resolution wind fields, NWPSAF-KN-TR-002, Version 1.2, 31-08-2006

World Meteorological Organisation, ″Statement of Guidance Regarding How Well Satellite and *In-Situ* Sensor Capabilities Meet WMO User Requirements in Several Application Areas″. WMO satellite reports SAT-26. *WMO/TD No 1052*, 2001.