CGMS-40 WMO-WP-19
V1, 2 October 2012

Prepared by WMO
Agenda Item: IV.1
Discussed in: Plenary

OPTIMAL DISTRIBUTION OF POLAR-ORBITING SOUNDING MISSIONS

The Impact of Temporal Spacing of Observations on Analysis Accuracy

Submitted by Dr John Eyre and Peter Weston, United Kingdom

This paper presents a theoretical study of the impact of the temporal spacing of observations on the average analysis errors in a simple system analogous to a numerical weather prediction system. The results are relevant to questions concerning the optimal distribution of polar-orbiting satellites, and particularly to the question of how available satellite assets might be deployed in the three orbital planes recommended by the WMO "Vision for the Global Observing System in 2025".

The results of this study show that the average analysis errors are indeed sensitive to observation spacing. Moreover, although the sensitivity is small when forecast error variances double at their average rate (~12 hours), it is much greater when doubling times are shorter (6 or 3 hours), as might be expected in some high-impact weather events. The results support the case for deploying satellites in roughly equally spaced orbits where possible.

Actions proposed:

Optimal Distribution of Polar-Orbiting Sounding Missions

The Impact of Temporal Spacing of Observations on Analysis Accuracy

1. INTRODUCTION

The work described in this paper was motivated by some simple questions concerning the orbital spacing of polar-orbiting satellites. Firstly, for a given quantity of observational assets, does the way in which these assets are distributed amongst different times of day make any difference to NWP analysis accuracy, and hence to forecast accuracy? Secondly, if it does make a difference, is it significant? Equivalent questions will arise in the design of other observing systems.

The intuitive or "common sense" answer to the first question, which has guided the design of most observing systems, is that observations should be spaced equally in time (and space), all other things being equal. Of course, all other things are not usually equal; ground-based observing systems have evolved with very unequal distributions in space, reflecting the relatively high costs of making observations over the oceans and in unpopulated areas, and also the disparities in economic resources between countries. Conversely, at the locations at which ground-based observations are made, they are usually made more uniformly in time, although, even here, unequal spacing of observations can arise for economic reasons, e.g. higher staff costs during local night-time.

The spatio-temporal coverage characteristics of space-based observations are controlled to a large degree by the characteristics of satellite orbits. Geostationary satellites allow observations to be made equally spaced in time and fairly uniformly in space within the coverage area of each satellite (which is determined by its longitude). The low Earth-orbiting satellites that support operational meteorology are usually placed in near-polar, sun-synchronous orbits, such that their instruments observe at fixed local times of day. Scanning instruments on such satellites can then provide almost complete global coverage twice a day (for thermal infra-red and microwave instruments) with almost uniform spatial resolution. When considered as a system, a number of such satellites will then deliver coverage at times determined by the satellites' local equator crossing times (ECTs).

The WMO "Vision for the Global Observing System (GOS) in 2025" (WMO, 2009) recommends a system of polar-orbiting satellites in three orbital planes, roughly 60 degrees apart, thus providing observational coverage approximately every 4 hours. (This applies at mid and low latitudes; coverage is more frequent at high latitudes where successive orbits give overlapping coverage, but in this paper we shall not consider further this nuance.) The justification for having at least three operational polar-orbiting satellites, rather than two, has been supported by NWP impact studies (e.g. Eyre and English, 2008). There have been fewer studies of the benefits of having three satellites in well-spaced orbits, rather than in any other spacing; we are only aware on one such study (Di Tomaso and Bormann, 2011), which showed the positive impact of microwave sounding observations from the combination of Metop-A, NOAA-18 and NOAA-15 satellites compared with those from the less widely spaced combination of Metop-A, NOAA-18 and NOAA-19. This work was reported to CGMS-38 (EUMETSAT, 2010).

In this paper, we firstly show that the "common sense" approach to observing systems design, i.e. that equally spaced in time is best, is supported by theory but not as simply as one might expect. In fact, it depends crucially on the metric that is used to quantify the performance of the system. Secondly, for some performance metrics, the degree to which temporal spacing of observations makes a difference to performance depends strongly on the rate of forecast error growth within the NWP system, which in turn is related to the rate of amplification of perturbations within the real atmosphere.

2. APPROACH

In this study, we examine the properties of a very simple data assimilation system, i.e. the evolution in time of the estimate of the error in a system with a single variable in space, when assimilating observations of the same variable. The observations are inserted with a repeating temporal pattern, but not necessarily equally spaced in time, simulating the observations from polar-orbiting satellites in up to three orbital planes.

Within this system, we compute the mean value of the analysis error variance over the repeat cycle of the observations, which in this study is 12 hours. We also compute the mean value of the "analysis accuracy", which is defined here as the inverse of the analysis error variance. The theory on which these calculations are based is described in Annex A.

The variables of this system are:

- the number of observations inserted in a 12-hour cycle – we use here either 1, 2, 3 or 4 observations, to simulate 1, 2, 3, or 4 satellites;
- the spacing of the observations within each 12-hour window – again, we choose spacings to simulate satellites in three orbital planes, as described in section 3;
- the observation error variance – we use a nominal value of 1.0, and so all calculations of analysis error variance should be considered as relative to this value;
- the forecast error growth rate, to which the results are very sensitive – discussed in more detail below.

Simmons and Hollingsworth (2002) discuss the error growth rate within a global NWP system. Updating these results, Simmons (2012) shows that the average doubling time for the standard deviation (SD) of error in the ECMWF operational global NWP system has been ~1.2 days in recent years, and that this value has decreased significantly compared with previous decades. Therefore it is realistic to assume that, in a few years time, global NWP systems will have doubling times of ~1 day for SD of forecast error, and so ~12 hours for error variance. We take 12 hours as the baseline value for the doubling time for forecast error variance in this study. However, we also note that many situations of high-impact weather occur when error growth rates are significantly higher than their average values. We therefore explore the consequences of shorter doubling times for forecast error variance: 6 hours and 3 hours.

3. EXPERIMENTS AND RESULTS

A set of experiments has been conducted computing analysis error variances and accuracies, and also their mean quantities over a sequence of time steps. In each experiment, a repeating pattern of observations is simulated with the error variance of each observation set to 1.0. See Annex A for further details of the method.

In each experiment, observations are inserted at 4-hour intervals, to simulate the potential of satellites in 3 equally spaced orbital planes. The experiments differ in the total number of satellites simulated and the way in which they are distributed between the three planes, as described in Table 1. Each experiment is assigned a "constellation code" to summarise how the observations are distributed in time.

relative observation time (hours) →			**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
experiment number	**number of observations**	**constellation code**												
1	1	[1,0,0]	1											
2	2	[2,0,0]	2											
3	2	[1,1,0]	1				1							
4	3	[3,0,0]	3											
5	3	[2,1,0]	2				1							
6	3	[1,2,0]	1				2							
7	3	[1,1,1]	1				1				1			
8	4	[2,2,0]	2				2							
9	4	[1,2,1]	1				2				1			

Table 1: Observation spacings for each experiment: values show the number of observations assimilated at each observation time.

	experiment number	1	2	3	4	5	6	7	8	9
	number of observations	1	2	2	3	3	3	3	4	4
Δt	constellation code	[1,0,0]	[2,0,0]	[1,1,0]	[3,0,0]	[2,1,0]	[1,2,0]	[1,1,1]	[2,2,0]	[1,2,1]
12 h	mean accuracy	1.485	2.970	2.970	4.454	4.454	4.454	4.454	5.939	5.939
	mean error variance	0.701	0.350	0.341	0.234	0.228	0.228	0.225	0.171	0.169
	% difference	-	+2.7	0	+3.6	+1.3	+1.1	0	+0.7	0
6 h	mean accuracy	0.764	1.528	1.528	2.291	2.291	2.291	2.291	3.055	3.055
	mean error variance	1.531	0.766	0.690	0.510	0.468	0.462	0.444	0.345	0.336
	% difference	-	+10.9	0	+15.0	+5.5	+4.1	0	+2.9	0
3 h	mean accuracy	0.404	0.808	0.808	1.212	1.212	1.212	1.212	1.616	1.616
	mean error variance	4.509	2.254	1.546	1.503	1.108	1.018	0.882	0.773	0.680
	% difference	-	+45.8	0	+70.5	+25.7	+15.5	0	+13.7	0

*Table 2: Mean analysis accuracy and mean analysis error variance for each experiment and for different doubling times of forecast error variance (**Δ**t).*

Table 2 gives the mean analysis accuracy and mean analysis error variance for each experiment. It also gives the percentage increase in mean analysis error variance compared with the experiment giving the lowest value for this quantity when the same number of observations is assimilated. The data in Table 2 is shown in graphical form in the following figures. Fig.1 and Fig.2 show the mean analysis error variances for all experiments, for forecast error variance doubling times of 12 hours and 3 hours respectively.

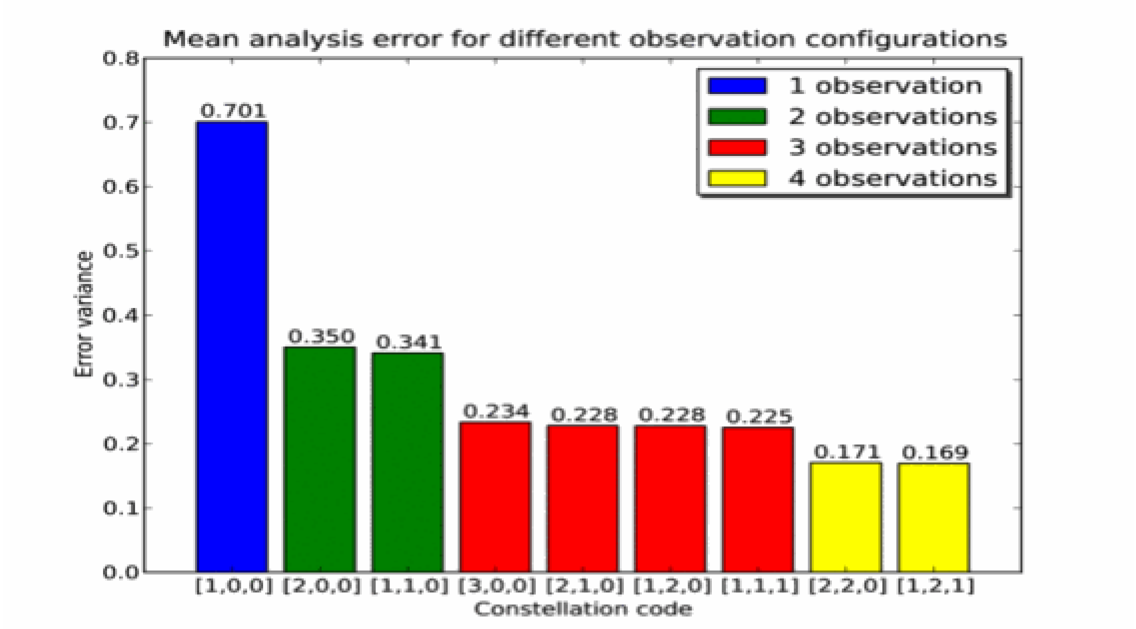

Fig.1: Mean analysis error variance for different observation configurations, with a doubling time for forecast error variance of 12 hours.

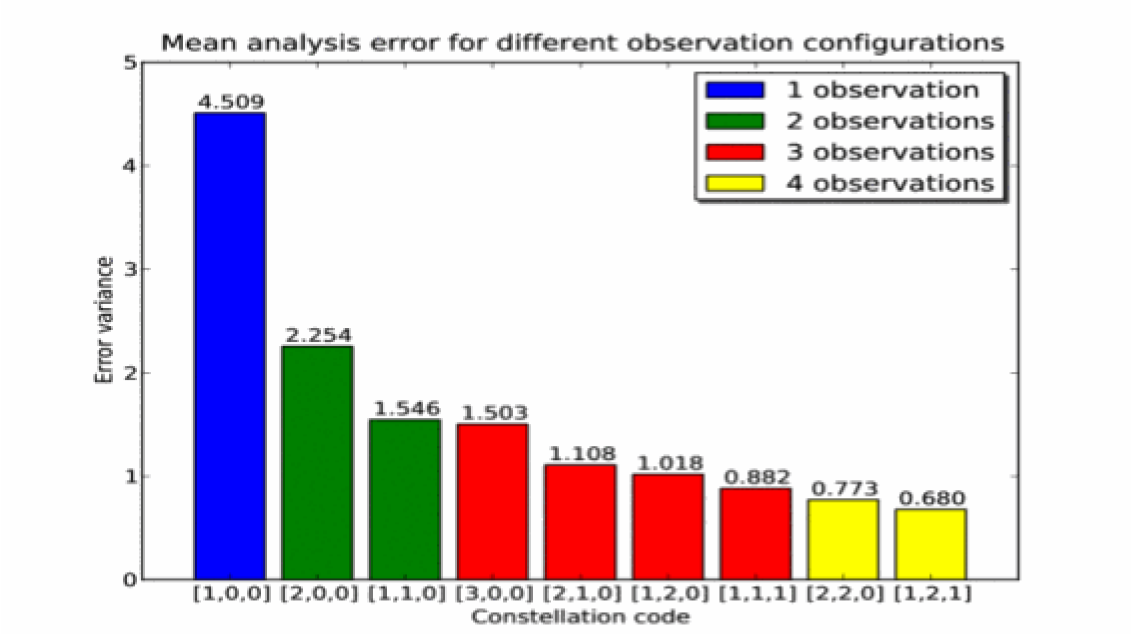

Fig.2: As Fig.1, with a doubling time for forecast error variance of 3 hours.

Fig.3 shows, for all forecast error growth rates and for experiments using 3 observations in each 12-hour window, the percentage increase in analysis error variance for each experiment relative to the optimal result which, in this case, is the [1,1,1] constellation.

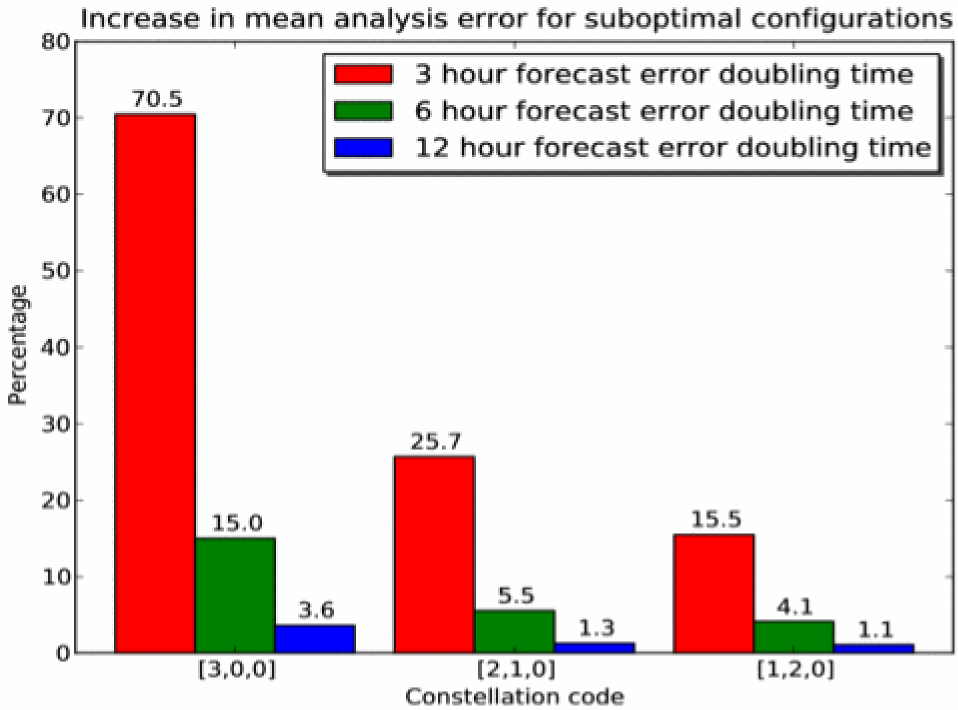

Fig.3: Percentage increases in mean analysis error variance for different doubling times of forecast error variance, for different configurations of three satellites relative to the best (i.e. the [1,1,1] configuration).

4. DISCUSSION

The results presented in Section 3 confirm the theoretical results derived in Annex A: that the mean analysis accuracy increases in proportion to the number of observations but is <u>independent</u> of how the observations are inserted within the 12-hour repeating pattern. Conversely, the mean analysis error variance <u>is</u> sensitive to the observation spacing, giving the lowest value for the most even spacing in time. Moreover, the increase in error variance caused by unequal spacing in time is very sensitive to the rate of forecast error growth. For example, for an error variance doubling time of 3 hours, the mean analysis error is ~70% higher when 3 (independent) observations are inserted at the same time compared with when they are spaced optimally (equally) in time.

Which of these results are most relevant to the design of a combined system of polar orbiting satellites? Firstly, are mean quantities most relevant (rather than the maximum or minimum values in the time series from which the means are derived)? Secondly, are metrics of analysis accuracy or analysis error variance most relevant? To answer the first question, we recall that, although sun-synchronous satellites provide observations at the same local time each day, the timing of the observation relative to the analysis/forecast cycle times for a global NWP system will vary depending on which part of the Earth is being observed. It is therefore not possible to optimise the observation times (relative to NWP cycle times) to benefit all parts of the world simultaneously. For this reason we argue that it is analysis accuracies/errors averaged over the observation repeat period that are most relevant.

To address the second question, we suggest that priorities of the application need to be considered; the main purpose of reducing errors in analyses is to reduce errors in subsequent forecasts and, in general, we are more interested in reducing errors in forecasts for which errors are large, rather than reducing errors that are already small. This suggests that we should focus attention on a metric that minimises the maximum analysis error, and this is equivalent to optimising the mean analysis error variance. (Conversely, if we use the mean analysis accuracy as a metric, this will tend to maximise the maximum accuracy within the cycle, i.e. to minimise error where it is already lowest.) Noting also the answer to the first question – that we need to consider effects on forecast errors in all parts of the world – we conclude that the most appropriate metric to use for this problem is the mean analysis error variance.

Let us now consider the results in Table 2 and Figures 1 to 3 that are most relevant to decisions over the next few years to optimise the polar orbiting satellite constellation. The nominal constellation consistent with the WMO Vision calls for satellites in 3 equally spaced orbits, i.e. constellation [1,1,1]. The WMO Vision also draws attention to the merits of multiple satellites in each orbit, but mainly for the purpose of ensuring operational back-up and robustness. Europe, via EUMETSAT, plans to ensure coverage in one orbit (~21:30 LECT), and the USA in another orbit (~13:30 LECT). Together, these assets represent the constellation [1,1,0]. China's plans include assets in both these orbits which, together with the American and European assets, would give rise to the constellation [2,2,0]. However, there is the possibility that China might move one of its future satellites into the otherwise empty orbit (~17:30 LECT), giving rise to the constellation [1,2,1]. (Note that, if we assume that all satellites carry equivalent observing systems, then constellations [1,2,1], [2,1,1] and [1,1,2] are all equivalent when measured in terms of mean analysis error variance.) This indicates the constellations that are most interesting to consider, and particularly how much [1,1,1] improves on [2,1,0] or [1,2,0], and how much [1,2,1] improves on [2,2,0].

For the 3-satellite constellations, [1,1,1] is significantly superior in mean analysis error to [1,2,0] or [2,1,0] even for an error variance doubling time of 12 hours (i.e. by >1%). For the shorter error doubling times of 6 and 3 hours, this advantage rises to 4-5% or 15-25% respectively. It is also interesting to note that [1,2,0] is superior to [2,1,0].

For the 4-satellite constellations, [1,2,1] is superior to [2,2,0]: by ~1% for a doubling time of 12 hours and by ~3% and ~14% at the shorter doubling times. For the reasons given in section 2, we suggest that significant weight should be given to the results for these shorter doubling times, as they are likely to be relevant for some high-impact weather events.

To what extent are these simple theoretical experiments (i.e. zero-dimensional in space) relevant to a real global NWP system (i.e. three-dimensional in space, with a wide variety of observations assimilated)? Recent results from the Met Office global NWP system (Joo *et al*., 2012) show that, for short-range forecasts, 64% of the forecast error reduction can be attributed to satellite observations and 36% to surface-based observations. Moreover, of the satellite data impact, ~90% comes from polar-orbiting satellites. (Comparable calculations by most other global NWP centres give similar results.) This suggests that, in a real global NWP system, the results would be attenuated a little because of the influence of other observations (surface-based and geostationary) but not greatly, and particularly not in the case of mid-latitude storms developing over the oceans where observations from polar orbit are dominant. Furthermore, although in this study the analysis-forecast system is highly simplified in space, it is likely to capture the essence of the impact of “blocks” of polar-orbiting satellites data which give coverage of a limited swath of the Earth at well-spaced intervals in time. We therefore maintain that this very simple study is well suited to capture the main features expected in the real-world problem.

5. CONCLUSIONS

The findings of this study may be summarised as follows:

The mean analysis error variance is dependent on the temporal spacing of observations, but the mean analysis “accuracy” (i.e. the inverse of error variance) is independent.

The mean analysis error variance is the most relevant metric when assessing the impact of the temporal spacing of observations on global NWP performance.

The dependence of mean analysis error variance on observation spacing is very sensitive to the assumed rate of forecast error growth:

- for a 12-hour doubling time of forecast error variance, the dependence on observation spacing is significant but small, for a set of observations simulating 3 or 4 polar-orbiting satellites in up to 3 orbital planes;
- for a 6-hour or 3-hour doubling time, the dependence is much greater; for a 3-hour doubling time reaching ~25% increase in variance for plausible 3-satellite constellations, and ~8% for 4-satellite constellations.

Because of the large contribution of satellite sounding data to forecast error reduction compared with other elements of the GOS, the simple experiments presented here are expected to be relevant to real NWP systems, and particularly for rapidly developing storms over mid-latitude oceans.

These results support the assumption guiding the WMO “Vision for the GOS in 2025” that polar-orbiting satellites should be equally space in time, as far as is practicable.

References

Di Tomasi E. and N.Bormann, 2011. Assimilation of ATOVS radiances at ECMWF. EUMETSAT/ECMWF Fellowship Progress Report No.22.

EUMETSAT, 2010. Results from studies supporting polar platform orbit considerations. CGMS-38 EUM-WP-41.

Eyre J.R. and S.J. English, 2008. Impact studies with satellite data at the Met Office. Proceedings of 4th WMO Workshop on "The impact of various observing systems on NWP"; Geneva; 19-21 May 2008.
http://www.wmo.int/pages/prog/www/OSY/Reports/NWP-4_Geneva2008_index.html

Joo S., J. Eyre and R. Marriott, 2012. The impact of Metop and other satellite data within the Met Office global NWP system using an adjoint-based sensitivity method. Met Office, UK; Forecasting Research Technical Report No.562. (Also submitted to Mon. Wea. Rev.)

Lorenz, E.N., 1982. Atmospheric predictability experiments with a large numerical model. *Tellus*, **34,** 505–513.

Simmons A.J., 2012. Assimilation of observational data for atmospheric monitoring and forecasting. Vilhelm Bjerknes Medal Lecture, EGU General Assembly; Vienna; 22-27 April 2012.

Simmons A.J. and A. Hollingsworth, 2002. Some aspects of the improvement in the skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **128**, 647-677.

Simmons A.J., R. Mureau and T. Petroliagis, 1995. Error growth and predictability estimates for the ECMWF forecasting system. *Q. J. R. Meteorol. Soc.*, **121,** 1739–1771.

WMO, 2009. The Vision for the GOS in 2025.
http://www.wmo.int/pages/prog/www/OSY/gos-vision.html

Annex 1

THEORY

We shall use the following terminology: A, B, R and F will denote the variance of error in the analysis, background, observation and forecast respectively. The inverses of these quantities – A^{-1}, B^{-1}, R^{-1} and F^{-1} - will be called “accuracies”.

We consider a data assimilation system in which observations are potentially available and assimilated at a series of time steps i. Using linear estimation theory, the analysis accuracy is given by:

$$A^{-1}_i = B^{-1}_i + R^{-1}_i . \qquad (1)$$

Note that, if there is no observation at time i, then $A^{-1}_i = B^{-1}_i$.

The background state at time i is obtained by forecasting the analysis at time i-1 forward in time. The amplification of forecast error variance over the time interval is assumed to be β:

$$B_i = \beta A_{i-1} . \qquad (2)$$

Combining (1) and (2),

$$A^{-1}_i = \beta^{-1} A^{-1}_{i-1} + R^{-1}_i . \qquad (3)$$

The mean analysis accuracy over a series of N time steps, from i=j+1 to i=j+N, is therefore:

$$(1/N) \Sigma_{i=j+1}^{j+N} A^{-1}_i = \beta^{-1} (1/N) \Sigma_{i=j}^{j+N-1} A^{-1}_i + (1/N) \Sigma_{i=j+1}^{j+N} R^{-1}_i . \qquad (4)$$

We assume now that the observational pattern in time steps j+1 to j+N is a repeating pattern, i.e. observations with the same errors were inserted in time steps j-N+1 to j, and j-2N+1 to j-N, etc. As the series is iterated, it converges to an equilibrium pattern, i.e. a repeating pattern of analysis values in which $A^{-1}_j = A^{-1}_{j+N}$. This allows us to replace the 2nd term in eq.(4):

$$(1/N) \Sigma_{i=j+1}^{j+N} A^{-1}_i = \beta^{-1} (1/N) \Sigma_{i=j+1}^{j+N} A^{-1}_i + (1/N) \Sigma_{i=j+1}^{j+N} R^{-1}_i , \qquad (5)$$

or $$(1/N) \Sigma_{i=j+1}^{j+N} A^{-1}_i = (1 - \beta^{-1})^{-1} (1/N) \Sigma_{i=j+1}^{j+N} R^{-1}_i . \qquad (6)$$

At first sight, eq.(5) represents a surprising result: it says that the mean analysis accuracy is determined only by the forecast error growth parameter, β, and the sum of the observation accuracies; it does not depend on how the observations are distributed over the time window between steps j+1 and J+N. This result can be understood by considering the combined effects of the terms in eq.(3); an “accuracy contribution” from each observation in the pattern is inserted into the analysis via the first term, and it then decays in the subsequent N time steps via the second term. Because the system is in equilibrium, the two effects are in balance, for each observation in the pattern, irrespective of where they are inserted in the cycle.

On the other hand, if we construct an equation similar to eq.(4) but for the mean analysis error variance, (1/N) $\Sigma_{i=j+1}^{j+N}$ A_i, we do not find a similarly simple term involving only sums of observation error variances. Consequently, we find that the mean analysis error variance does depend on how the observations are distributed in time. We also find that the lowest retrieval error variance is found when the observations are distributed uniformly in time. This can be proved analytically for some simple cases and it can be demonstrated numerically for these and more complex cases, as will be shown in Section 3 of this paper.

We consider now the forecast error growth parameter, β, in more detail. Several authors (e.g Lorenz, 1982; Simmons et al, 1995, and their references) have considered mechanisms of forecast error growth and appropriate parametric representations for it. Here we assume a form that is appropriate in the early stage of the forecast when the errors are small and far from saturation. In this limit, we can assume:

$$dF/dt = \alpha F . \qquad (7)$$

Integration of this equation leads to:

$$\alpha = \ln 2 / \Delta t , \qquad (8)$$

where Δt is the doubling time for forecast error variance. Combining eq.(8) with eq.(2) we obtain:

$$\beta = \exp(\ln 2 \, \delta t / \Delta t) , \qquad (9)$$

where δt is the time step in equations (1)-(6).

It is important to note that the doubling time, Δt, is for error variance whereas in much of the literature the doubling time is calculated for standard deviations (SDs) errors in quantities such as height or pressure. The two are simply related: the doubling time for error variance of a quantity is half that of the doubling time for the SD of the quantity.

Using the theory presented above, a set of experiments has been conducted computing analysis error variances and accuracies, and also their mean quantities over a sequence of time steps, in the following way. In each experiment, a repeating pattern of observations is simulated with the error of each observation set to 1.0. The analysis accuracy is calculated through eq.(3), which is iterated to convergence, i.e. until the pattern of analysis accuracy over a period N time steps repeats itself. For the observation patterns and error growth rates used in this study, a 7-day integration period is found sufficient to achieve convergence. A time step of 1 hour is used, and observations are inserted on 12-hour repeating patterns. Average analysis accuracies and error covariances at convergence are computed for the last 12 hours of this 7-day period.

When two or more observations are assimilated at the same time step, it is assumed that their accuracies are additive on eq.(1) and eqs.(3)-(6). This is equivalent to assuming that their errors are uncorrelated. For the assimilation of satellite sounder radiances into state-of-the-art NWP systems, this is a good approximation, provided that the observations are not so close in time and space that NWP model "errors of representativeness" are significant. It is therefore assumed that observations are appropriately filtered in space and time to avoid this problem.