The impact of satellite data on NWP

Anthony McNally

European Centre for Medium-range Weather Forecasts (ECMWF)
Shinfield Park, Reading RG2 9AX, UK
Anthony.McNally@ecmwf.int

ABSTRACT

This study looks at the impact of satellite observations on the quality of NWP forecasts. The various factors that can influence impact are described - and it is shown that many of these are not directly related to the intrinsic quality of the measurements. A critical examination is made of two approaches that are widely used to evaluate the impact of observations: The first are Observing System Experiments (or **OSE**) where data are deliberately denied from the assimilation and forecasts from this depleted system compared to those of a control system (containing all observations). The second approach is Adjoint Sensitivity Diagnostics (**ASD**) where the impact of all observations is evaluated within a single control system by comparing forecasts from the background and analysis states. It can be argued that both of these approaches have their respective strengths and weaknesses. Using the two approaches the current impact of the major satellite observing systems is evaluated. Microwave and infrared sounding clearly emerge as the most important drivers of headline forecast skill - although other satellite measurements make important contributions to the forecasting of other parameters at short lead times.

1 Introduction

Modern data assimilation systems are highly complex and can extract atmospheric information from a wide variety of different observations. It is important to regularly assess and understand the relative contribution of each component of the observing network to the overall health of the forecasting system because:

- The impact of observations may change over time depending on model / DA evolution and the availability of new data.
- It is important to explore resilience and redundancy to optimise the use of resources (computational and human diagnostic).
- It is useful for the long term planning of the global observing system - not just deciding which instruments fly, but in which orbit and when.

More recently the quantification of observation impact has become a necessity to help justify the enormous costs of developing, building and launching satellite instruments. Of course this requires an additional (non-scientific) translation of forecast impact to socio-economic benefits, but this is beyond the scope of this particular study.

2 What factors influence forecast impact?

It is very important to understand the factors that influence the forecast impact of satellite observations, in particular to distinguish between those factors that are intrinsic to the measurement (e.g. the quality of the data) and those which are not (e.g. how and where the data are used).

Obviously careful consideration of the latter can avoid making incorrect decisions about the usefulness (or otherwise) of an observation that might be quite different in a different set of circumstances.

Measurement quality: In this context we refer to how accurately an instrument measures the quantity that is actually being observed. Satellites observe radiation at the top of the atmosphere and poor (noisy) detectors, poor radiometric calibration and spectral calibration will result in bad data. Poor data can of course still have a limited positive impact if the assimilation scheme accurately characterizes its error characteristics, but significant systematic errors are very difficult to handle.

Measured quantity: However accurately an instrument measures radiation at the top of the atmosphere - it may still be very difficult to extract useful atmospheric information from these data. A radiance observation that is sensitive to many atmospheric phenomena and therefore requires great skill on the part of the assimilation scheme may have a smaller impact than an observation that is much simpler to use. An obvious example is that of extracting temperature information from infrared radiances (that are highly sensitive to clouds and atmospheric composition) which is much harder than extracting the same information from microwave radiances (which are much less sensitive to interfering phenomena). It is also important to consider the measured quantity in terms of what is already well known in the assimilation system. An observation that brings new or unique information is likely to have a larger impact than one that measures a quantity already measured by a myriad of other sensors.

Spatial coverage and time: Clearly an observing system that has limited spatial coverage (either due to the orbit/scanning of the instrument or because large amounts of the data are unusable e.g. due to clouds) will have a limited impact on the forecasting system. However, increasingly NWP centres now employ four dimensional assimilation schemes where the time that the observations are made may have a strong influence upon their impact. In a 12 hour assimilation window, observations made near the end of the window potentially have a stronger influence than those near the beginning. This is because the forecast model can evolve numerous atmospheric variables over time to fit the data at the end of the window.

Tuning of the assimilation system: Poorly specified observation or background errors may have a dramatic influence upon the impact of an observation. These quantities essentially combine to determine the weight given to the data. Too little weight and the observation will have no influence at all - too much weight and the assimilation scheme may begin to interpret random measurement noise in the data as genuine signal. While the former may result in the data being underused, the latter can easily lead to a perfectly good observation having a negative impact on forecast quality.

3 Evaluating forecast impact

Observing System Experiments are a long established method for determining the impact of an observation. Forecasts are run from a control assimilation that makes use of all available observations and these are compared against forecasts run from an assimilation that has the observation of interest deliberately withheld. The reduced forecast errors in the control system with respect to the denial system can reasonably be attributed to the presence of the additional observation and is thus a measure of its impact. The advantage of OSE is that they tell us very directly and exactly what we wish to know – namely what happens to forecast errors with / without a particular observation. We may look at the impact upon different forecast parameters, in different ranges and in different areas. The obvious disadvantage is that OSE are very expensive to run - particularly if we wish to test the impact of many different observation types individually over long periods (this is often necessary to obtain statistically robust results). A less obvious issue with OSE is that in denying an observation we may change the error characteristics of the system and the original specification of background errors may not be valid (such as the observations that are retained may not be used optimally).

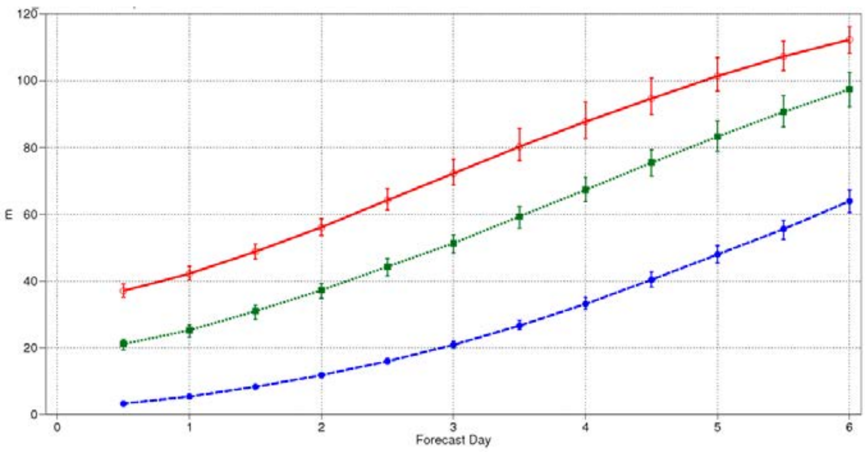

Figure 1: Mean forecast errors for 500 hPa geopotential height in the Southern Hemisphere for a control system (blue) and two NOSAT systems (where all satellite data are removed). In red the background errors are not retuned to take account of the satellite data denial. In green a crude attempt is made inflating the background errors by a factor of two.

An example of this is shown in Figure 1 where all satellite data are denied in an OSE. Compared to the control a massive inflation of forecast error occurs when all of the satellite data are removed (in excess of three days of skill at day 5). However, even with a very crude attempt to retune background errors (in this case just multiplying the background errors of the control by a factor of two) the NOSAT system can recover a significant amount of skill. By inflating the background errors we give a much more appropriate weight to the retained observations (in this case in situ data) which then have a larger corrective influence on the assimilation system. Obviously this OSE is a rather extreme example where effectively 90% of the observations have been removed and a retune of the background errors is absolutely essential. For smaller denials (e.g. where just a single sensor is removed) background errors are unlikely to change as dramatically, but we should at least be aware that if we do not retune the system we will generally overestimate the impact of the observation.

Adjoint Sensitivity Diagnostics have additionally been applied to assessing the forecast impact of observations (Baker and Daley, 2000; Gelaro et al., 2007; Cardinali 2009). Exact implementations vary at different NWP centres, but essentially these methods compare the errors in short-range forecasts launched from either end of the assimilation window (i.e. from the analysis and background states). It is argued that the reduced error in (e.g. 24hr) forecasts from the analysis compared to the error in the (e.g. 36hr) forecast from the background is due to the work done by the observations to improve the description of the atmospheric state. Using the adjoint of the forecast model the forecast error differences (usually expressed as a total energy norm) are traced back in time (linearly) and attributed to individual observations according to their weight in the analysis and relevance to the chosen metric. The advantage of ASD is that they are very inexpensive (certainly compared to OSE) and do not require the assimilation system to be perturbed (that may render assumptions about background errors invalid). The obvious disadvantage of ASD is that they can only operate in the very short-range (typically 24 hours) and can tell us nothing about how an observation impacts forecasts in the medium range (around 5 days).

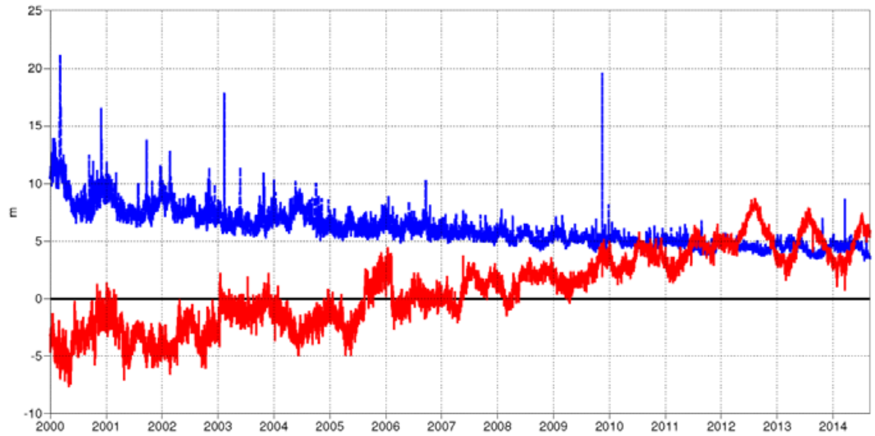

Figure 2: A time series of random and mean differences between independent atmospheric analyses of 500 hPa geopotential height from the UK Met Office and ECMWF.

An issue that affects both OSE and ASD is that of *verification* - namely what state of the atmosphere do we consider *truth* and compare forecasts against to measure forecast error. Conventional or in situ observations provide reasonably accurate and direct estimates, but the spatial coverage is poor and biased towards populated areas and flight routes. Satellite observations provide excellent spatial coverage, but measure radiance and not detailed vertical profiles of atmospheric parameters. These considerations have led to the widespread use of NWP analyses to verify forecasts and to understand the limitations of this we need to understand how accurate NWP analyses are. By attributing the various contributions to observation innovation statistics authors such as Simmons and Hollingsworth (2001) estimated analysis errors for 500 hPa height to be in the region of 5 m to 10 m. If we compare the analyses from two independent NWP centres (as equally valid estimates of the truth) we obtain random differences that are at the lower end of this range. However, it can be seen in Figure 2 that seasonally varying mean errors would put the total RMS uncertainty in our atmospheric analyses firmly in the range 5 m to 10 m.

If we examine the RMS forecast errors for the control plotted in Figure 1 (blue line) we see that while medium range errors are typically around 50-60 m, values for the very short range (24hrs) are in the range 5 m to 10 m. Thus uncertainties in verifying analyses are exactly the same magnitude as the forecast errors we are trying to measure. This means that our evaluation of short-range forecast errors is significantly less reliable than medium range forecast errors and this must be taken into consideration when drawing conclusions from OSE and ASD (particularly the latter as it can only be applied in the short range).

4 An assessment of current satellite impact on NWP forecasts

A set of OSE have been performed to evaluate the impact of major components of the satellite observing system. The experiments use the latest version of the ECMWF operational forecasting system (CY40R1) with 137 vertical levels (surface to 0.01 hPa) and a reduced horizontal resolution (T511 ~ 40 km). The period tested ran from 1st March 2014 to 30th June 2014 and the observations considered are shown in Table 1.

All conventional (in situ) data	**CONV**	TEMP/AIRCRAFT/SYNOP/ SHIP
All satellite data (all groups below)	**SAT**	
LEO microwave sounding radiances	**MWS**	7 x AMSUA, 1 x ATMS, 4 x MHS
LEO infrared sounding radiances	**IRS**	2 x IASI, 1 x AIRS, 1 x HIRS
GEO data (AMVs and radiances)	**GEO**	2 x GOES, 2 x METEOSAT, 1 x MTSAT, polar AMVs
GPS-RO bending angle data	**GPS**	COSMIC, 2 x METOP-GRAS
Microwave imager radiances	**MWI**	1 x TMI, 1 x SSM/IS
Scatterometer surface wind data	**SCAT**	2 x ASCAT

Table 1: Observation groups considered in the OSE study

The control run assimilates all of the above observations and provides the standard for a set of experiments where each of the groups listed in Table 1 are deliberately withheld. Only the experiment where ALL satellite observations are removed uses modified background errors (the simple inflated values described previously). Other denial experiments use the same background errors as the control. Forecast errors are evaluated using the ECMWF operational analyses as verification.

The collective impact of satellite observations compared to that of conventional (in situ) data is shown in Figure 3.

As expected the loss of all satellite data causes a catastrophic loss of forecast skill in the Southern Hemisphere (equivalent to two days of skill at day-5) whereas the impact of losing in situ observations is barely discernible. In the Northern Hemisphere the in situ data clearly have more of an influence (particularly in the very short range), but the overall forecast skill is still dominated by the collective impact of satellites.

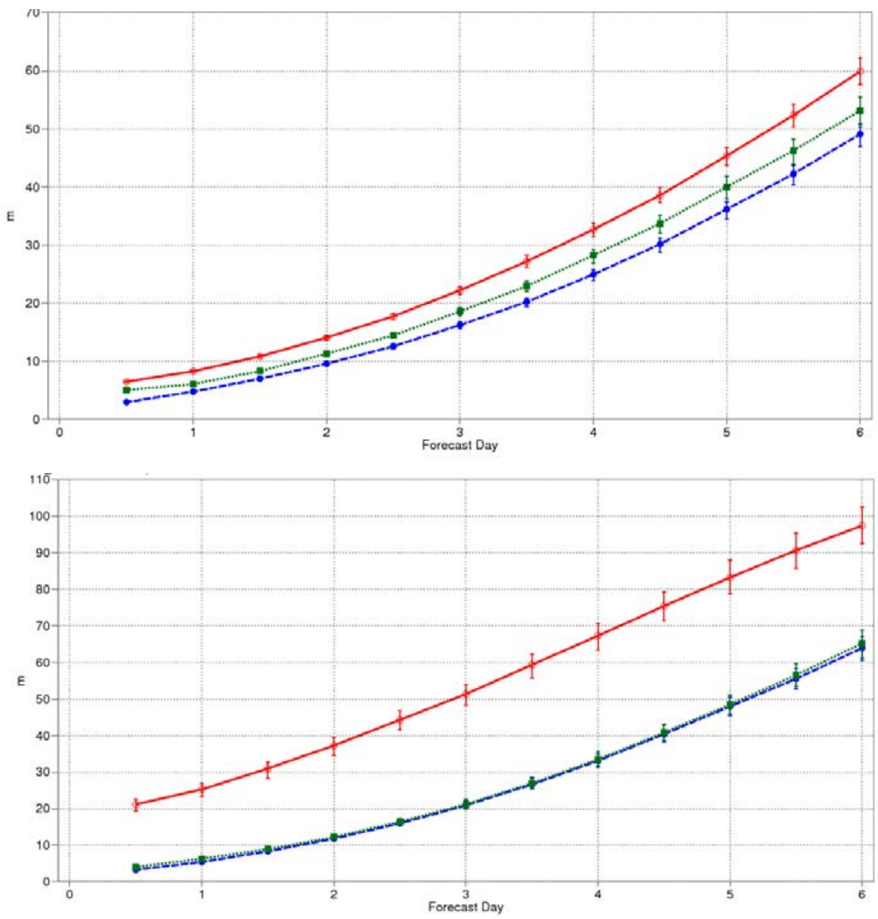

Figure 3: RMS forecast errors for 500 hPa geopotential height averaged over the test period in the Northern Hemisphere (upper panel) and Southern Hemisphere (lower panel) for the control system (blue), NOSAT system (red) and NO-CONV system (green). Vertical error bars indicate 95% statistical significance intervals.

We now look at the impact of individual components of the satellite network, but to display these results in a concise manner forecast errors are differenced and normalised with the respect to the errors of the control system. In Figure 4 the ranked normalised results for 24 hour forecasts are shown, where it can be seen that in both hemispheres, collectively the in situ data are more important than any of the individual satellite components (the highest ranking of which is microwave sounding). Indeed the magnitude of the loss of skill when single satellite components are withheld is rather small compared to the collective impact of the satellites (from Figure 3). This would suggest that there is a high degree of resilience in the satellite network such that if one is component is lost, others compensate to maintain the skill of the system.

It is only at the 24 hour range that we can attempt to compare ranking of individual components from the OSE with the assessment from the Adjoint Sensitivity Diagnostics (shown in the Figure 5). In the Northern Hemisphere the ASD results largely agree with the OSE, ranking in situ data highest with microwave and infrared sounding the most important satellite contributors. However, in the Southern Hemisphere ASD ranks the contribution of in situ data much lower than that suggested by the OSE (demoting from top rank to third).

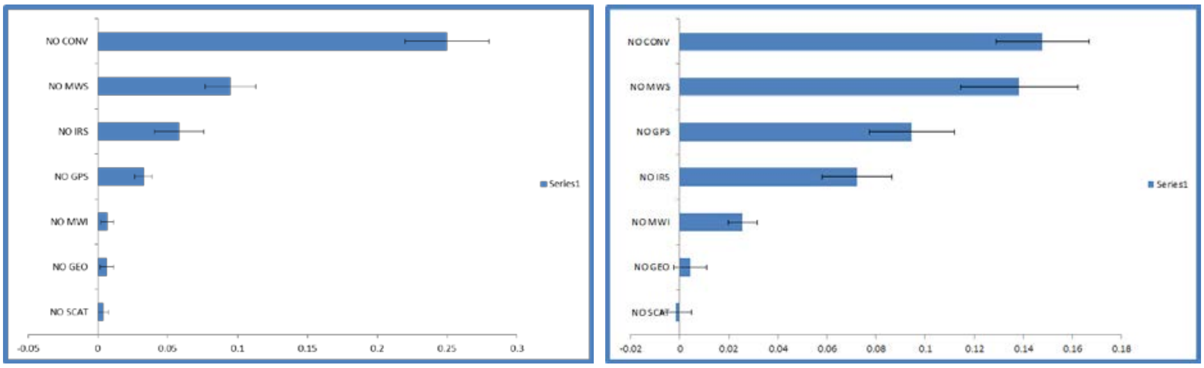

Figure 4: RMS forecast errors at 24 hours for each of the OSE denial experiments (differenced and normalised with respect to the forecast error of the control) for 500 hPa geopotential height averaged over the test period in the Northern Hemisphere (left panel) and Southern Hemisphere (right panel). Horizontal error bars indicate 95% statistical significance intervals.

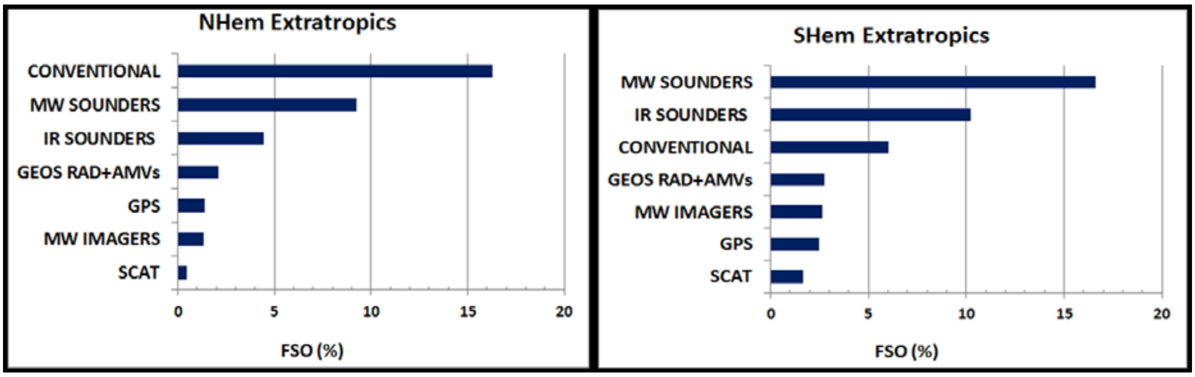

Figure 5: Ranked contributions to forecast error reduction (evaluated for a total energy norm) estimated by Adjoint Sensitivity Diagnostics averaged over the test period in the Northern Hemisphere (left panel) and Southern Hemisphere (right panel)

How much significance we attach to this disagreement about the importance of in situ data in the Southern Hemisphere is not clear given the previously discussed uncertainties in the verifying analysis. However, the OSE and ASD do agree that in the short range MWS is the dominant satellite component and in situ data are clearly still very important.

We now turn to examine results from medium range forecasts for which we only have results from the OSE. In the Northern Hemisphere it is possibly remarkable that in situ data still dominate over any single component of the satellite network. Of the satellite components microwave and infrared sounding are again clearly the most important, but overall the magnitude of the skill lost due to individual satellite data denials is even less in the medium range compared to the short range.

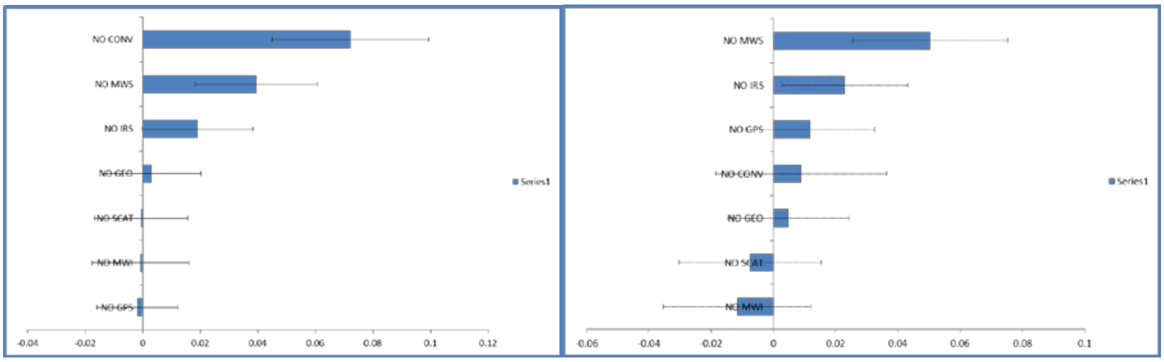

Figure 6: RMS forecast errors at 144 hours for each of the OSE denial experiments (differenced and normalised with respect to the forecast error of the control) for 500 hPa geopotential height averaged over the test period in the Northern Hemisphere (left panel) and Southern Hemisphere (right panel). Horizontal error bars indicate 95% statistical significance intervals.

This again reinforces the idea that the satellite network is resilient to the loss of individual components (even significant ones) and it seems that the degree to which satellites can compensate for the loss of others is even greater in the medium range.

So far our attention has been on headline scores of 500 hPa geopotential height in the Northern and Southern Hemispheres where microwave and infrared sounding data dominate. However if we look at other parameters in other areas we start to see impacts of other satellite components.

Radiances from microwave imagers are primarily sensitive to boundary layer humidity (once the effects of clouds and precipitation have been accounted for). Results from the OSE confirm the importance of these data to constrain humidity structures near the surface. However, it is also clear (in Figure 7) that this impact is rather short lived as MWI drop from the top rank at 24 hours to fourth by day 6.

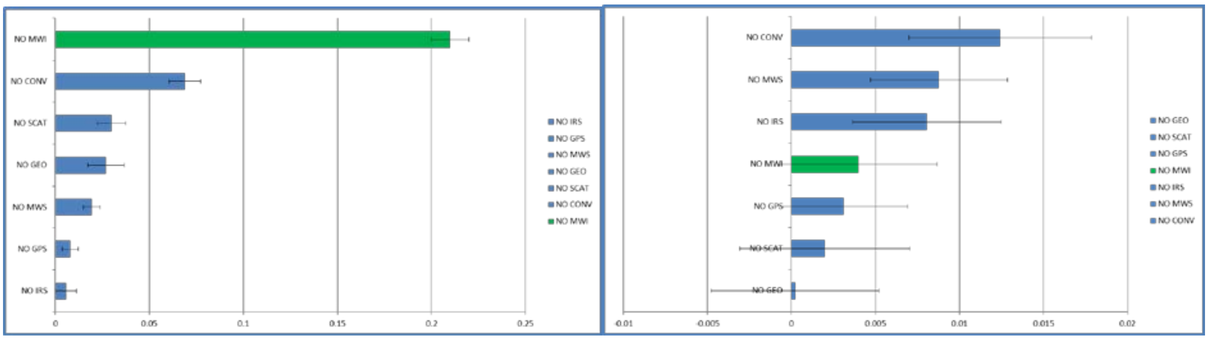

Figure 7: RMS forecast errors for each of the OSE denial experiments (differenced and normalised with respect to the forecast error of the control) for 850 hPa relative humidity averaged over the test period in the Tropics. The left panel shows results at 24 hours and the right panel results at 144 hours. Horizontal error bars indicate 95% statistical significance intervals.

Data from geostationary satellites are primarily used to constrain the wind field either by the direct assimilation of Atmospheric Motion Vectors (AMV) or by dynamical tracing of humidity sensitive radiances in the upper troposphere. Previous studies have shown that these wind data are most useful in the Tropics where it is more difficult to exploit balance considerations and extract wind information from thermal data. Wind scores from the OSE are shown in Figure 8 and confirm the strong impact of the GEO data, but once again we see the influence is short lived. GEO data drop from the top ranked observations at 24 hours to the bottom rank by day 6.

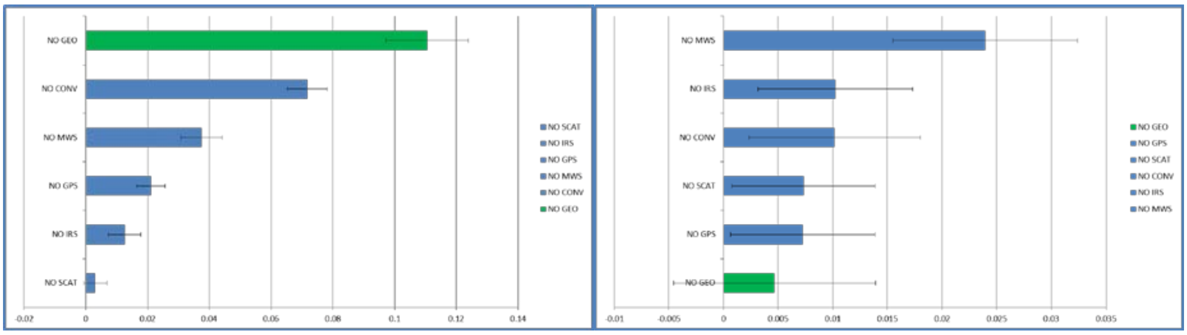

Figure 8: RMS forecast errors for each of the OSE denial experiments (differenced and normalised with respect to the forecast error of the control) for 200 hPa vector wind averaged over the test period in the Tropics. The left panel shows results at 24 hours and the right panel results at 144 hours. Horizontal error bars indicate 95% statistical significance intervals.

Finally the impact of GPS-RO data is mainly seen in forecasts of lower stratospheric temperatures. It is clear in Figure 9 that the influence persists throughout the forecast range, but furthermore, that GPS is mainly limiting the growth of systematic (mean) forecast errors (included in the RMS figures but not the standard deviation). This strong constraint upon mean errors is expected as GPS are the satellite observations that are not bias corrected in the data assimilation system (essentially against the model).

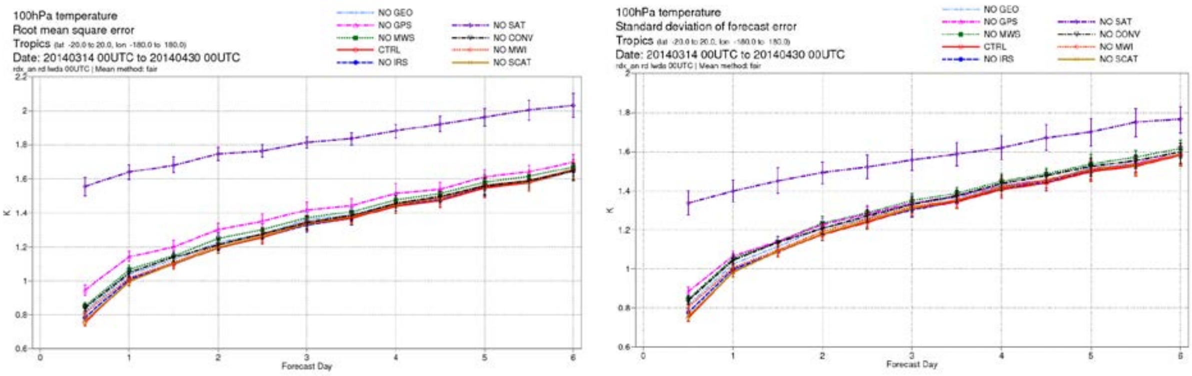

Figure 9: RMS (left panel) and standard deviation (right panel) of forecast errors for 100 hPa temperature averaged over the test period in the Tropics. The OSE denial of GPS data is shown in pink, others as detailed in the legend. Vertical error bars indicate 95% statistical significance intervals.

5 Summary of results and conclusion

Many factors influence the impact a particular satellite observation may have on reducing forecast errors. These of course include the quality of the measurement and what is measured, but also extend to when and where the observations are taken, their ease of use and the correct tuning of key parameters such as background and observation errors.

Furthermore, methods to measure forecast impact are prone to error and misinterpretation and we must be very careful. In particular uncertainties in our knowledge of the true state makes a reliable verification of short range forecast rather difficult.

Some clear features emerge from this latest study of forecast impact:

- The simultaneous loss of all satellite data results in a catastrophic loss of forecast skill in both hemispheres and is much more damaging than the removal of all in situ (conventional) observations.
- However, there is high degree of resilience in the satellite network such that when just one component is removed the others compensate and there is only a modest loss of forecast skill (compared to removing all satellites).
- Individual OSE denials indicate that microwave and infrared sounding radiances are the main drivers of headline forecast skill, but other sensors (GEO and MWI) have impacts on specific parameters (wind and boundary layer humidity respectively) in the short range (but reliable verification is an issue). However, in the medium range MW and IR sounding dominate the forecasting of even these parameters presumably by constraining the larger scales.
- GPS-RO observations have a strong constraining effect limiting the growth of systematic temperature errors in the lower stratosphere. They are the only observations assimilated with no bias correction and thus represent a crucial anchor on mean errors.
- Not discussed in this paper (but highlighted at the seminar talk) is the fact that satellite data can have a large impact on forecast quality when they are used offline for diagnosis and improvement of models. In the case of both low level humidity and upper level winds in the Tropics it is interesting to note how microwave and infrared sounding data replace the more direct observing systems (that dominate the short range) at the top ranked positions in to the medium range.

Acknowledgements

Many thanks to Cristina Lupu who provided the input material on ASD.

References

Baker, N. and R. Daley, 2000: Observation and background adjoint sensitivity in the adaptive observation-targeting problem. *Q. J. R. Meteorol. Soc.*, **126**, 1431–1454.

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.*, **135**, 239–250.

Gelaro, R. and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179–193.