The direct assimilation of cloud-affected infrared radiances in the ECMWF 4D-Var

Tony McNally

ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, UK
Tony.McNally@ecmwf.int

1. Introduction

There are three major incentives to be able to use cloud-affected infrared radiance observations.

- The restriction of using only infrared sounding data that can be pre-determined as clear represents a major under-exploitation of very high cost instruments such as the Atmospheric Infrared Sounder (AIRS) and the Infrared Atmospheric Sounding Interferometer (IASI). Estimates of cloud cover vary in the literature, but any instruments with a footprint of order 10 km will typically only yield between 10–30% completely clear soundings.
- Only using radiances in clear-sky has the potential to bias the assimilation system towards particular synoptic or climatological regions (e.g. areas with low humidity) which may not be representative of the wider atmospheric conditions.
- There is evidence to suggest that cloudy areas are meteorologically sensitive and that constraining analysis errors in these regions (with observations) is important to limit forecast error growth.

The observed infrared radiance spectra from satellites contain potentially useful information on clouds in addition to the temperature and humidity information they were designed to provide. Cloud parameters have been estimated simultaneously with temperature and humidity in offline retrieval schemes. More recently it has been demonstrated that for weakly affected radiances there is potential to pass externally generated cloud parameters from a 1D-Var retrieval as fixed values to a global direct radiance assimilation scheme. There are significant scientific aspects (discussed in the next section) in which the estimation of cloud parameters differs from the estimation of temperature and humidity. These issues have been addressed in the standalone retrieval algorithms, but as yet have precluded the extension of global operational assimilation schemes such as 4D-Var to estimate cloud parameters from the direct assimilation of cloud-affected infrared radiance data.

2. Particular issues related to cloud analysis

The assimilation of radiance data (or indeed any observation) may be regarded as a correction process where errors in a background estimate of the atmospheric state are corrected by comparison with the observations. In variational schemes such as 4D-Var, temperature errors in the rather accurate short-range forecast background typically translate into a radiance signal of just a few tenths of a degree Kelvin – with radiance observation errors being of a similar magnitude. Signals are slightly larger in the case of humidity, but still typically of order 1 to 2 degrees Kelvin. In contrast, the presence of clouds can easily introduce signals of order many tens of degrees Kelvin in infrared radiance observations and we can see that there is a considerable mismatch between the magnitude of temperature (and humidity) information in the observations compared to the cloud information. While this does not present any theoretical obstacle to the simultaneous estimation of clouds, it does demonstrate the need to be extremely careful when handling the cloud information so as not to destroy the important temperature and humidity information.

Data assimilation schemes require a degree of linearity between the observations and the analysis variables – at least for small perturbations (the so called tangent linear approximation). Put more simply, this requires that the relationship that links the observed quantity (radiances in the case of satellites) to the parameters we wish to estimate in the analysis (e.g. temperature and humidity) should not vary dramatically in different atmospheric conditions. For radiance measured in clear-sky conditions this is reasonably valid. For humidity sensitive radiances the approximation is weaker, but still reasonable. However, in a cloudy sky the dependency on the atmospheric state (i.e. the actual cloud conditions) is potentially extreme. This extreme behaviour means that to exploit the tangent linear approximation within the analysis a very accurate initial description of the cloud conditions is required.

Another potential complication for the inclusion of clouds within the analysis is the availability of background information. For temperature and humidity it has already been stated that the short-range forecast provides rather accurate background information and we have fairly reliable estimates of the error covariance describing this accuracy. Modern NWP models do provide highly detailed cloud information, but it is probably fair to say that opinion is divided as to its accuracy. Recent satellite missions such as CLOUDSAT will certainly improve our understanding and validation of NWP cloud information in the future, but as yet we do not have an accurate quantification or errors in NWP model clouds.

A further issue relates to matching cloud information provided by the model to the observations. For variables such as temperature a simple interpolation to the observation is sufficient, but for clouds this is potentially more problematic (particularly if there are variable model resolutions involved in the analysis – not representative of the spatial scale of the satellite).

Finally we must consider the issue of how to accurately model the cloud contribution in the infrared radiance data. The most sophisticated and accurate cloudy radiative transfer schemes require inputs of cloud liquid and ice content profiles, cloud fraction profiles and even microphysical parameters that describe the detailed radiative characteristics. In addition, important assumptions must be made about how fractional clouds at each level overlap in the vertical. The next section describes the steps that have been taken to overcome these difficulties.

3. Extending the 4D-Var analysis for clouds

The analysis has been extended to estimate parameters that describe cloud – namely the cloud top pressure (CP) and the effective cloud fraction (CF). The latter takes into account semi-transparency and represents the equivalent amount of opaque cloud in the instrument field of view. The two extra cloud variables do not exist as spatially continuous fields (as would be the case for example with model fields of temperature or humidity), but as local variables defined and estimated only at satellite observation locations.

The framework used was originally designed to allow the simultaneous analysis of surface parameters from satellite data inside the 4D-Var. It allows different observations with their own particular spatial scales and spectral characteristics to produce independent estimates of quantities such as skin temperature and emissivity – where a single model field would be unlikely to be representative of what the satellite actually observes (e.g. a single NWP model field of skin temperature could not be simultaneously appropriate for microwave and infrared data where the surface penetration depth is very different). In this sense the framework is very suitable for cloud variables where values of the parameters may be strongly dependent on the characteristics of the satellite (e.g. size of the footprint).

Cloud analysis variables handled in this framework are not correlated spatially and explicitly coupled with any other analysis variables through the background error statistics or the 4D-Var strong model constraint. They are also not transported forward in time to the next assimilation window. However, the estimation of the new cloud variables is strongly constrained by the coupling of clouds and existing analysis variables (such as temperature and humidity) in the radiative transfer simulation of the observed radiances. Background estimates of the additional analysis variables that describe cloud are not taken from the NWP model for the reasons discussed previously. A pragmatic approach has been adopted that uses a small subset of the available radiance channels at a given location to estimate a first guess of the cloud parameters (essentially a least squares fit of CP and CF to the observations).

4. Selection of overcast data

An important and unique aspect of this scheme is that cloud-affected radiances are only used when the scene is determined to be completely overcast. The justification for this is that many of the problems of analyzing cloud are alleviated when the scene is overcast: difficulties related to the forward modelling of multi-layer clouds and how fractional clouds at different levels overlap in the vertical are removed. The highly simplified cloud analysis variables can be regarded as a good representation of the real cloud conditions. Interactions between surface variables (e.g. skin temperature) and cloud variables in the analysis are removed. Overcast clouds (particularly in the mid to upper troposphere) are most accurately determined by the radiance observations and we may have a high degree of confidence in the accuracy of the background estimate. Finally, we have the possibility of obtaining very high vertical resolution temperature information at the overcast cloud top. Here, many lower-tropospheric and surface sensing radiance channels that would normally (i.e. in clear-sky) provide deep layer temperature information are suddenly only sensitive to a very thin atmospheric layer just above the cloud and together provide a very accurate estimate of its temperature.

Not all overcast scenes are used in the current system. These include cases where the least squares algorithm estimates clouds that are un-physically high, clouds below the surface or clouds with a fraction greater than one (note that no physical bounds are placed as prior constraints upon the algorithm). Overcast scenes when the cloud top is determined to be below 900 hPa are also excluded. While these are quite plausible and physically reasonable, the confidence we may have in the accuracy of low clouds is less and the presence of certain synoptic conditions (such as sharp inversions associated with marine stratocumulus clouds) have proved problematic for convergence in the analysis. Finally, overcast data over land and ice surfaces are not used. While the treatment of overcast data should be insensitive to the underlying surface (as the surface is essentially obscured), a poor prior knowledge of the surface skin temperature or emissivity may cause large errors in the initial cloud parameters and lead to an erroneous identification of a location being overcast.

If the scene is determined as overcast in the initial estimation process the cloud amount is fixed at 1.0. This is obviously sub-optimal as some of the additional information that becomes available in the main analysis may suggest a reduction in the cloud fraction. However, it is a pragmatic measure avoiding the complications of fractional cloud cover. Having fixed the cloud amount, all available channels are then activated in the analysis. The only exceptions are the short-wave channels of HIRS and AIRS which are considered unsuitable as they may have a significantly different cloud sensitivity compared to the long-wave channels (note that no short-wave channels from IASI are currently used at ECMWF).

If the scene is not overcast (or is overcast, but fails to meet the above criteria) the cloud fraction (CF) is fixed at zero and the system reverts to a clear-sky treatment of radiance data – the same as that of the baseline assimilation. The cloud detection will attempt to find a subset of channels (if any) that are unaffected by the cloud and only these will be activated in the analysis.

5. Overcast data coverage

An example of the combined additional coverage provided by the use of overcast data from the four infrared sensors is shown in Figure 1 for a typical 12-hour analysis window. For reference the clear-sky data usage is also shown. Overall, the extra overcast data available to the analysis only amounts to approximately 10% of the total. This low yield is mainly due to the stringent quality control criteria that are applied – principally the exclusion of very low clouds (which actually account for the majority of overcast scenes). Nonetheless, it can be seen that the cloudy data that survive the quality control still provide a good filling of gaps in areas where clear-sky radiances are not available. Features such as large frontal regions and the storm tracks are conspicuous, but there is a fairly even distribution of extra data to be found in all areas of the globe.

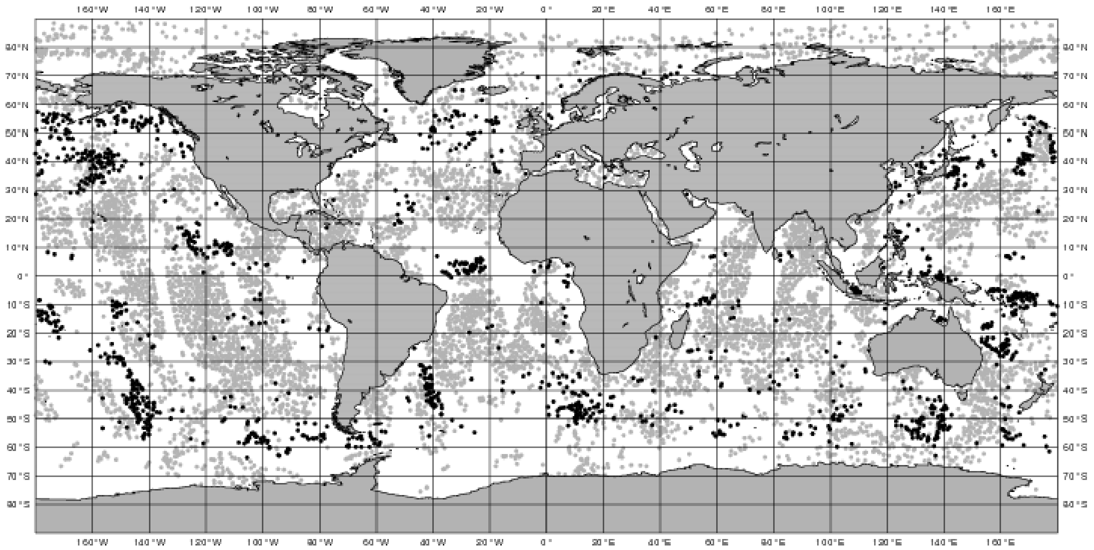

Figure 1: The distribution of overcast cloudy scenes (black dots) and completely clear scenes (grey dots) from the four infrared sensors METOP-HIRS, NOAA-18 HIRS, METOP-IASI and AQUA-AIRS for the first 12 hour analysis cycle 00 UTC on 12 January 2008. {cloudy-clear-window-cov.ps}

6. Analysis increments from overcast radiances

The impact of the overcast infrared radiance data upon the assimilation system has been investigated by examining the difference between analysis increments from a system with clear and overcast data compared to a control system that only uses clear-sky data (shown in Figure 2 for temperature at two different levels 250 hPa and 700 hPa). As expected the increments from the two systems differ in exactly the locations where additional overcast radiance observations are available. Furthermore there is a good correspondence between the altitude where the changes occur and the diagnosed height of the overcast cloud. Above very high clouds the temperature changes are in the upper troposphere and there are no changes lower down. Above low clouds there are temperature changes in the lower troposphere, but these can also be accompanied by differences at upper levels.

The correspondence between cloud altitude and the impact of the overcast radiances is better illustrated in Figure 3 where the vertical profile of analysis increments are shown for two locations in the equatorial Pacific (just east of Papua New Guinea). At the first location (Figure 3a) only overcast HIRS radiances are used (above clouds diagnosed at altitudes between 225 and 240 hPa) and at the second location (Figure 3b) only overcast IASI radiances are used (above clouds diagnosed at altitudes between 195 and 215 hPa). In both cases modified temperature changes at the cloud top are evident, but appear sharper in the case of IASI.

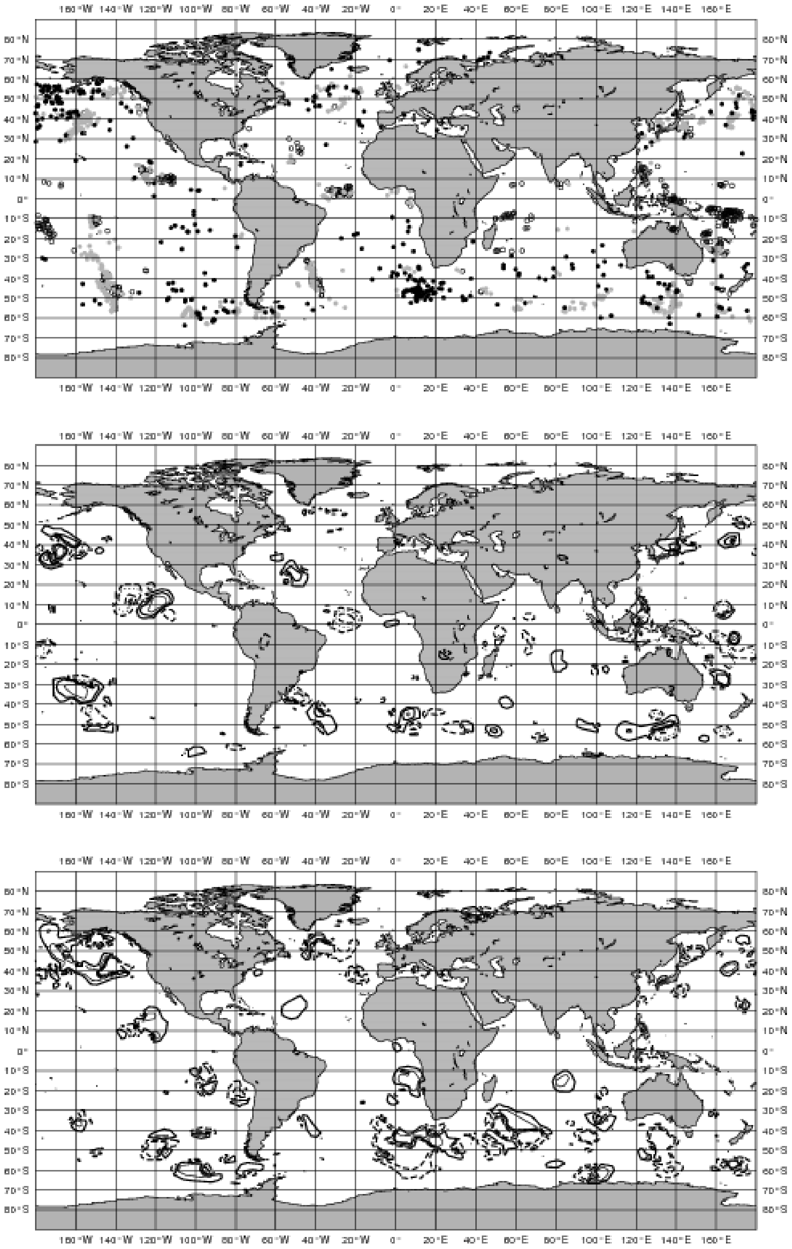

Figure 2: (a) Location of overcast scenes for the first 12-hour analysis cycle (00 UTC on 12 October 2008) separated into three categories depending on cloud height 850 hPa to 600 hPa in black, 600 hPa to 300 hPa in grey and 300 hPa to 100h Pa in open circles. (b) Temperature increment differences (EXPT minus CTRL) at 250 hPa. (c) Temperature increment differences at 700h Pa. For (b) and (c) the contours are at 0.2 K intervals – solid indicating positive values and dashed indicating negative values. {cover-and-inc.ps}

While the meteorological conditions at the two locations are not identical (e.g. the clouds at the HIRS location are diagnosed slightly lower than those at the IASI location) we cannot draw too strong a conclusion from this. However, it is a suggestion of what we expect to see – namely higher vertical resolution increments from the advanced infrared sounders compared to HIRS.

While these examples of increment changes demonstrate the ability of the overcast radiances to influence the analysis in a manner (and in regions) that would not be possible with clear-sky data, we need to verify if the increments cause the analysis to move closer to the true atmospheric state.

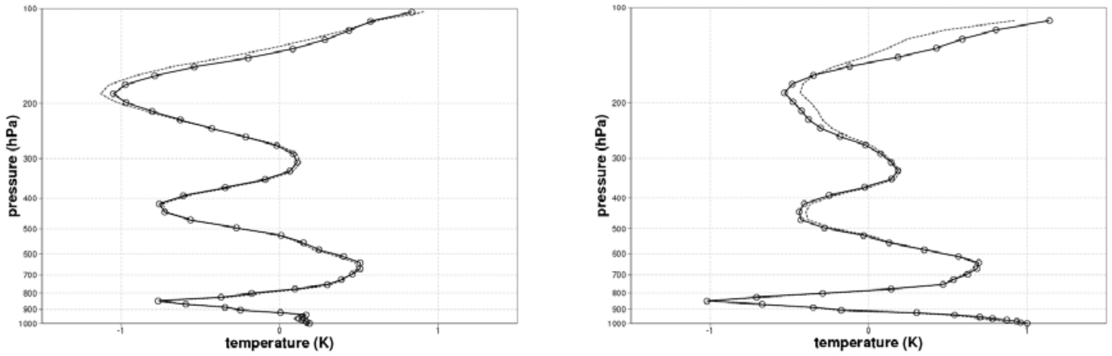

Figure 3: Vertical profiles of temperature increments from (a) HIRS overcast radiances, cloud altitudes diagnosed between 225 and 240 hPa and (b) IASI overcast radiances, cloud altitudes diagnosed between 195 and 215 hPa (lower panel) during the first 12-hour analysis cycle (00 UTC on 12 October 2008) from two tropical locations east of PNG. Solid lines indicate the CTRL and dashed lines indicate the cloudy EXPT system. The axes are temperature in Kelvin and pressure in hPa. {HIRS-high.ps}{IASI-high.ps}

Figure 4 shows temperature increment difference at 700 hPa caused by the assimilation of overcast radiances averaged over a one month period. The shaded areas indicate where the increments are smaller with the overcast data and open contours indicate the opposite. It can be seen that, in the vicinity of isolated oceanic radiosonde observations, the increments of the system using the cloudy data are reduced compared to the clear-sky control (e.g. in the North Atlantic and Pacific and also in the Southern Oceans near Bouvetoya, Crozet, Prince Edward and Kerguelen Islands). Reduced increments at these isolated radiosonde locations indicate that temperature errors are being better constrained in the surrounding oceans where satellite radiances are the dominant source of data. Thus we have some evidence (although the signal is small) that the additional use of overcast radiances is improving the quality of the assimilation with respect to independent observations.

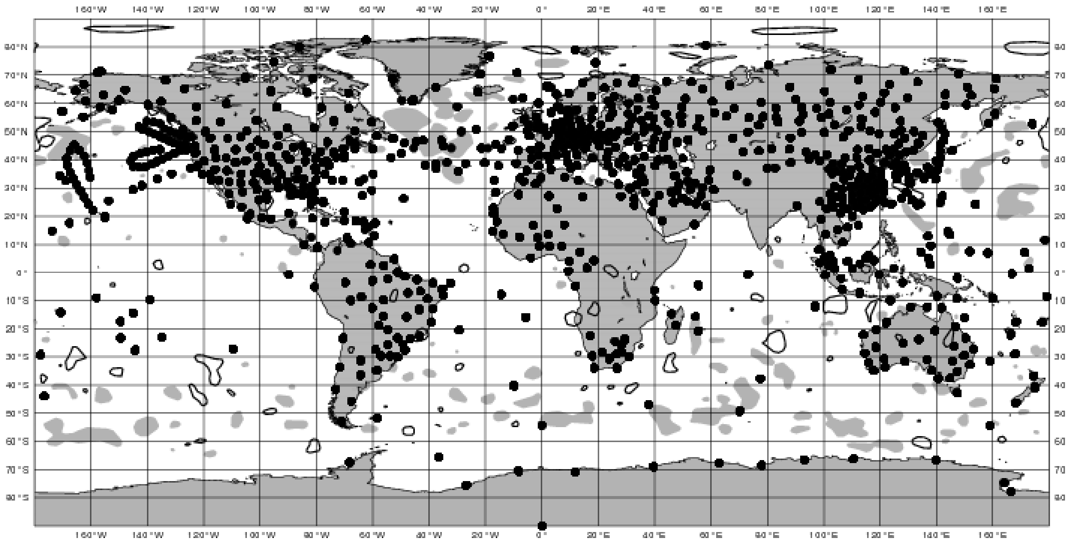

Figure 4: Root-mean-square temperature increment differences at 700 hPa (EXPT minus CTRL) averaged over the first month of the test period (12 January to 12 February 2008). Grey shaded contours indicate negative values i.e. where the EXPT has smaller RMS increments, open contour indicate the opposite (both contours begin at 0.1 K). Overlaid in black dots are the locations of radiosonde stations reporting and being assimilated at least once during the period. {T700inc-stat.ps}

7. Impact of overcast radiances on forecasts

The impact of using overcast radiance data on forecast quality has been tested over a three-month period. Figure 5 shows a comparison of normalized root-mean-square errors (cloudy system minus clear-sky system) for forecasts of 500 hPa geopotential height in the extratropics of the northern and southern hemispheres. Superimposed upon the plot are vertical error bars indicating the statistical significance of the differences evaluated with a standard t-test at the 95% level. In the northern hemisphere the two systems score the same. In the southern hemisphere the average forecast error is reduced using cloudy data, but the statistical significance of the differences is marginal beyond day one.

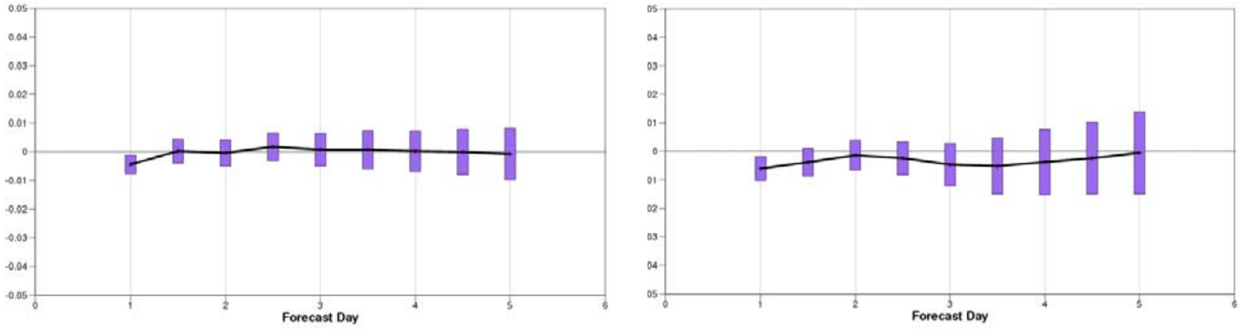

Figure 5: Normalized root-mean-square forecast error difference (EXPT minus CTRL) divided by CTRL for 500 hPa geopotential height in (a) northern hemisphere extratropics (20°–90°N) and (b) southern hemisphere extratropics (20°–90°S) evaluated over 77 cases of the full test period. Each system is verified against its own analyses and the vertical error bars indicate 95% confidence intervals. {NHz500.ps}{SHz500.ps}

The geographical distribution of day five temperature forecast error differences averaged over the same three-month period is shown in Figure 6. These demonstrate a positive impact of the overcast radiance data particularly at 200 hPa and 700 hPa in the North Atlantic and North Pacific. The magnitude of the changes in the tropics appears small in these maps, but normalized forecast error differences demonstrate a significant positive impact of the overcast radiances.

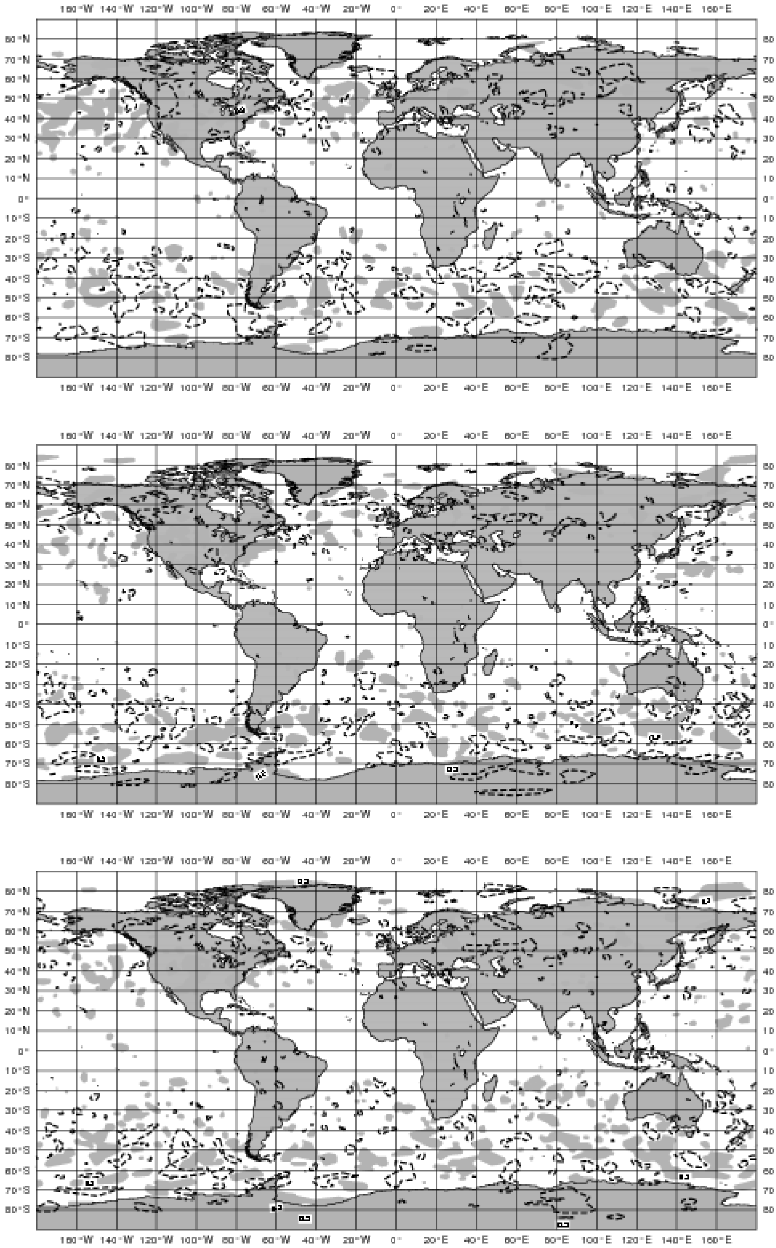

Figure 6: Maps of day five root-mean-square forecast error difference (EXPT minus CTRL) for temperature at (a) 200 hPa, (b) 500 hPa and (c) 700 hPa evaluated over 77 cases of the full test period. Each system is verified against its own analysis. Grey shaded contours indicate areas where the CTRL forecast error exceeds the EXPT by > 0.2 K, open contours indicate the opposite. {T-FCERR-maps.ps}

8. Summary and future work

The ECMWF 4D-Var analysis system has been adapted to directly assimilate cloud-affected infrared radiance observations in addition to the clear-sky data. The extension of the analysis control vector to include simplified cloud parameters allows the estimation of cloud variables simultaneously with all other atmospheric variables. The scheme only assimilates cloud-affected radiances in overcast conditions – thus avoiding many complications associated with the forward modelling and analysis of fractional cloud. In focussing on completely overcast scenes and using all available channels at these locations, this scheme departs from the more traditional approach of trying to use only weak or moderately cloud-affected radiances.

Assimilation experiments have been run with the new scheme in which overcast radiances from HIRS, AIRS and IASI are used in addition to the available clear-sky data. The overcast data locations typically represent 10% or less of the total due to the application of a stringent quality control. The extra data that are used give rise to modified increments (largest for temperature) at and above the diagnosed cloud top. The vertical scales of the temperature changes are much finer than would normally be obtained with clear-sky radiances. Information from independent radiosonde observations (particularly isolated oceanic stations) suggests that the quality of the assimilation in the mid-lower troposphere is slightly improved using overcast radiances compared to the system that uses only clear data. Forecasts are similarly improved when the overcast radiance data are used.

There are two important issues related to the longer-term future development of this scheme that should be discussed. Firstly, the strict limitation of using only completely overcast situations could be relaxed slightly and modest variations in cloud fraction allowed during the main 4D-Var analysis. This has not been tested so far as the initial emphasis has been placed upon constructing and testing a conservative and robust scheme for use in the operational ECMWF system. Relaxing the overcast limitation will certainly stretch the validity of the simplifying assumptions made here, but the trade off with a higher yield of data may well lead to a better overall system performance. This will be investigated in the near future. The possibility of treating significantly non-overcast situations in this simplified framework is probably rather limited. The description of fractional overlapping cloud would require many additional analysis variables and the estimation of these without reliable independent prior information on clouds (level by level) is likely to be prohibitively under-determined. However, one potential source of information linking the distribution of multilevel clouds is the physical parametrization of the NWP model.

This leads to the second important issue for future development, namely the explicit coupling of the estimated cloud parameters to other atmospheric variables. In its current form the system essentially uses cloud variables as a sink for cloud signal to enable the estimation of temperature and humidity. While the separation of cloud signals is strongly constrained in the main analysis by constraints imposed on temperature and humidity (note this is not the case in a standalone cloudy retrieval), the cloud information is then discarded. However, the knowledge of whether a scene is cloudy or not could (in addition) be digested by the assimilation scheme and lead to thermodynamically consistent adjustment of all atmospheric variables via the adjoint of the model physics. This has already been achieved in the operational use of rain affected microwave radiance data. Using this as a template, the very acute sensitivity of infrared radiances to cloud may prove to be an additional potent constraint upon the assimilation system in the future.