ECMWF/EUMETSAT NWP-SAF Workshop

Efficient Representation of Hyper-spectral Infrared Satellite Observations, 5 - 7 November 2013

Report Summarising Working Group Discussions (DRAFT)

The following organisations were represented in the working group discussions:
NWP_SAF, ECMWF, EUMETSAT, NOAA, METEO-France, CNES, DWD, UK Met Office, KIAPS (Korea),

Working Group 1: Fiona Smith (chair), Marco Matricardi, Pascal Brunel, Pascale Roquet, Hyoung-Wook Chun, Stephen Tjemkes, Christina Köpken-Watts, Peter Weston, Niels Bormann, Steve English, Cristina Lupu, Nadia Fourrié, Thomas August, Paul Poli, Elias Holm

Working Group 2: Andrew Collard (chair), Reima Eresmaa, Tim Hultberg, Vincent Guidard, Nigel Atkinson, Stefano Migliorini, Lothar Wolf, Rolf Stuhlmann, Ken Holmlund, Mark Parrington, Javier Andrey-Andres, Ed Pavelin, Adrien Deschamps, Julie Letertre-Danczak, Stephan Havermann

(R: indicates a specific recommendation)

Compression for Data Dissemination

The group was tasked with discussing possibilities for reducing the cost of disseminating high spectral resolution satellite data.

Near-lossless compression - a number of approaches were considered including the use of dedicated and proprietary software tools. It was felt that these probably could not provide the compression ratios that would be required to make a significant impact on cost. Also, some compression approaches were computationally demanding for encoding / decoding and could introduce significant delay to processing. During these investigations options for improvements within the current data format (BUFR) have been identified by optimising scaling factors to account for quantisation noise.

R: No near-lossless options have been identified that could achieve significant compression factors

Relaxed availability requirements - a significant driver on the current cost of dissemination is the stringent requirement that full resolution data is available to users in near-real-time, a very high percentage of the time (typically 95%). It was argued that considerable cost savings could be made if users could tolerate only having access to a reduced subset of the data (e.g. 300-400 channels) with very high availability (e.g. 95%) to support their operational services. Users could still have access to the full data stream (e.g. to support development activity), but this would be with a significantly relaxed timeliness and availability threshold. Indeed the latter could possibly be achieved not by dissemination, but by users having fast access to a rolling archive of recent data.

R: Data providers and real-time data users should investigate if this two-tier approach is a viable strategy (and if there are any implications for NWP-SAF AAPP products)

PCA compression for dissemination - EUMETSAT studies have shown that a carefully trained reduced set of principal component scores (e.g. 300) can convey a vast majority of the useful signal that was contained in the original observed spectrum. While not lossless - it is mostly noise that is lost and this could actually be an advantage to some applications (e.g. chemistry). Given the massive compression ratio provided by PCA - it was felt that it had to be considered as an option for data dissemination. The group expressed concerns about the stability of the data error characteristics when significant changes occurred in the atmosphere (e.g. sudden major volcanic events or slower decadal drifts) or when routine updates are made to the eigenvector training sample (e.g. to reflect changes in the noise characteristics of the instrument). It also questioned how quickly after the launch of a new satellite - stable PCA compressed data would be available to users (given that some time would be needed to establish a stable eigenvector training).

R: NWP centres (NWP_SAF) are encouraged to begin investigating the currently available near-real-time EUMETSAT IASI PC products to establish if these could be used operationally.

R: Similar investigations should take place within the chemistry community (Ozone SAF)

R: EUMETSAT Explore if the full instrument error covariance matrix is required for noise normalisation - is this stable and is detector-by-detector noise normalisation is needed

R: A mutually acceptable update strategy needs to be devised and documented for the EUMETSAT IASI PC products including distribution of revised principal components

Compression and Archiving

The group discussed if and how developments in data volume for real-time dissemination should impact the strategy and plans for archiving. The overwhelming consensus was that the archiving of full resolution level 0 and level 1 data must be continued. It was felt that nobody could anticipate how the data will be used in future decades for retrospective climate reanalysis and thus the original integrity of the data should be preserved. In addition to the long-term archive the possibility of EUMETSAT maintaining a fast access rolling archive was discussed - and how such a facility could alleviate pressure on the real-time dissemination resources to obtain research data.

R: Archiving of level-0 and level-1 data should be continued and be lossless - information on all metadata and processing should also be archived.

R: EUMETSAT should consult the user community on requirements for a more streamlined access to archived data - possibly exploring synergies with the ECMWF MARS

Timing of Key decisions

Given that the data dissemination strategy is a significant element when planning future satellite programs - the groups was asked to identify any key decisions and when they would have to be made. It was felt that there was still considerable uncertainty in what users could be willing to accept in terms of reduced / compressed data dissemination. The currently available options for volume reduction presented at the workshop needed to be investigated and tested by users with some urgency. However, it was acknowledged that telecommunication technology and indeed the demands of the user community can evolve rapidly and data providers should wherever possible build in a degree of flexibility to their plans. Key dates in satellite program planning were identified:

For MTG-IRS the level-1 data format must be provided to industry very soon and final decisions will be made by the end of 2014. For EPS-SG the full program proposal is anticipated in 2015.

R: Data users to evaluate possible data compression and reduction options and provide input to EUMETSAT as soon to assist program planning.

R: EUMETSAT and the NWP-SAF to make plans as flexible as possible in order to react to evolution in telecommunications technology and user requirements.

Compression for Data Assimilation

The group was asked to consider the options for exploiting data compression approaches for assimilation - with a view towards communicating information from the observed spectrum to the analysis in an efficient a manner as possible. While it was noted that compression for dissemination and assimilation are clearly linked at a practical level - the group discussions were largely based upon the assumption that science arguments rather than dissemination issues would not drive development of improved assimilation systems.

Possible efficient interfaces to the analysis - The assimilation of truncated principal component scores (**PCS**), reconstructed radiances (**RR**) and level-2 retrievals (**L2R**) were all discussed - and it a fundamental level if everything is done correctly - these were considered theoretically equivalent. The success or otherwise of any particular approach was thus thought to depend more on how well important elements of the assimilation system (discussed below) could be constructed and tuned in a practical operational environment.

Assimilation of PCS - A novel technology currently being developed and tested at ECMWF (in collaboration with EUMETSAT) where the measured spectrum is projected on to a fixed eigenvector basis (the synthetic eigenvectors of the PC-RTTOV radiative transfer model). The PCS are directly assimilated in the 4D-Var (using PC-RTTOV and its adjoint) and preliminary results suggest that the use of PCS in clear sky is comparable to the existing operational radiance assimilation system. The group felt it was important that ECMWF investigate if the PCS data generated operationally by EUMETSAT (as a solution to the dissemination problem) could be assimilated in this framework. This would require a re-projection from the truncated real data eigenvector basis to the synthetic (PCRTTOV) eigenvector basis, but otherwise no major obstacles were foreseen.

It was shown that in principle the same approaches to handling cloud in radiance space could be adopted in PCS space (i.e. using data above clouds and overcast scenes). However, to achieve this in practice appears technically demanding - requiring dedicated eigenvector projections to be computed on the fly to suit the particular cloud conditions. The handling of clouds could be more straightforward if the Jacobians of the PCS were more localised in the vertical. It was suggested that this could be achieved by imposing locality as a constraint upon the calculation of the projection - but in the discussions no details of how this might be done were proposed.

R: ECMWF investigate the assimilation of the IASI PCS data operationally disseminated by EUMETSAT

Assimilation of RR - Discussions concluded that this was arguably the simplest approach to implement for centres that already have a well-developed infrastructure for real radiance assimilation. The use of RR would not require any of the significant technical and scientific investment needed to develop a system to directly assimilate PCS. Technology developed for handling clouds in real radiances would also readily apply to RR. The group reviewed some (rather dated) ECMWF assimilation results and more recent work performed at the UK Met Office. It was clear that RR assimilation was certainly an attractive option, but it was pointed out that great care was needed in channel selection to respect the true information content of the data to avoid poor conditioning. The accurate specification of dedicated RR observation errors (including correlations) was also considered vital. It was questioned if the assimilation of RR actually required a dedicated observation operator that accurately simulated the multichannel nature of each individual the RR channel - which would be prohibitively expensive. This is still under investigation, but the consensus was that approximating the observation operator with the simple RTTOV calculation of the corresponding real channel should suffice as long as the inaccuracy is accounted for in the observation error covariance. Given the simplicity of implementation - it was felt that NWP centres and chemical assimilation users should begin serious testing of the RR approach, and at an appropriate stage an inter-comparison between the PCS, RR and real radiances should be carried out.

R: NWP centres and other users should begin investigating the use of RR and as soon as appropriate a clean comparison should be made between RR and PCS (and radiances)

Assimilation of L2R - It was thought that the assimilation of externally generated retrieved profiles (or components thereof) did not offer significant compression opportunities if there was a need to transmit Averaging Kernels (AK) with each observation. However, it was shown that using a L2R approach purely as an internal transformation to a reduced dimension space (within the data assimilation system) was at least theoretically viable. While no examples of exist of this being employed in an operational assimilation environment the group thought the approach warranted further investigation.

R: Studies be conducted in to the use of transformed retrievals within the data assimilation systems

Random and systematic error (and monitoring) - There was agreement that whatever interface to the analysis was employed, that an accurate modelling of the full observation error covariance would be required. There are well established techniques to do this and results from these methods applied to real radiance data and PCS data were shown at the workshop. It was suggested that these diagnostics should be applied to RR. Concerns were raised that the spectral non-locality implicit in both PCS and RR might result in error correlations that could allow errors from residual cloud contamination propagating through the entire atmospheric column. Similarly, stratospheric model errors could potentially impact the troposphere via the same mechanism. It was felt this was potentially a problem for NWP and chemistry and warranted very careful study.

So far the approach for the correction of biases has followed that for radiance data. The VarBC adaptive bias correction has been successfully applied to PCS and RR data with no obvious problems. It was argued that if RR display no mean residual with respect to the real radiances that no additional bias correction would actually be needed. Drifts have been observed in bias corrections

for PCS - and it was suggested that same processes that cause drifts in radiance biases (time varying model error and feedback with QC) could be responsible. Concerns were raised regarding the non-physical nature of PCS and RR. In the past - routine monitoring of radiance biases provided useful feedback on the forecast model, the RT model and instrument problems. If the community moved away from assimilating real radiance data would this valuable diagnostic information be lost? Overall the group felt that biases (in particular regional variations and time stability) needed to be studied further.

R: Users investigate the implications of spectral non-locality for the characterisation of observation error covariances for PCS and RR

R: Users and EUMETSAT study the time stability and regional variations of bias in PCS and RR.

Regional assimilation - No specific investigation of compression for data assimilation in the context of regional systems has been performed so far. The regional community is essentially following developments in global NWP. The group discussed the possibility of regional artefacts being present in globally generated PCS or RR data (both random and systematic) and agreed that these needed to be studied with some urgency. It was also pointed out that regional short-range forecasting puts greater emphasis on hydrometeor assimilation - and to date the preservation of information on clouds microphysics in truncated PCS and RR has not been studied. The group welcomed the initiation of a EUMETSAT funded fellowship at METEO-France (in 2014) to look at regional aspects of using compressed data.

R: EUMETSAT and regional NWP users should investigate if globally generated PCS or RR contain regional random or systematic artefacts

Reanalyses for climate studies - As is the case for regional applications, no specific investigation of compression for data assimilation in the context of climate reanalysis systems has been performed so far. The reanalysis community will follow developments in global NWP assimilation (indeed it will generally use state of the art global NWP systems). The primary requirement of the climate community are expressed in the discussions relating to archiving - namely that the full resolution uncompressed data (and metadata used in calibration) are archived. If the real-time NWP disseminated data moves towards a compressed format, provision should be made for intermittent reprocessing activity to support reanalysis. Scientifically the long term stability of PC compressed data is a concern - in particular drifts on decadal scales due to changing trace gas concentrations. It was felt that operational centres should begin routine monitoring of PC compressed data to investigate this.

R: Full resolution uncompressed data (and metadata used in calibration) must be archived irrespective of developments in real-time data dissemination.

R: EUMETSAT and NWP centres (via the NWP-SAF) should begin routine monitoring of compressed data products