Science Meeting 2 - observation error covariance studies

Meeting held on 27th April 2022

Background

The purpose of this meeting was to establish progress in observation error covariance estimation studies since the 2019 ECMWF/NWPSAF Workshop on the treatment of random and systematic errors

The working group were asked to review the following presentations from that workshop before the meeting:

- Overview by Niels Bormann
  - Slides: https://ecmwfevents.com/i/cd708733-df67-4adb-8572-64e7ad3d4262/oral-presentations/332dcc49-efbe-46ed-8afb-97a61211651c
  - Presentation video: https://vimeo.com/474651646
- Situation dependent errors by Alan Geer
  - Slides: https://ecmwfevents.com/i/cd708733-df67-4adb-8572-64e7ad3d4262/oral-presentations/06e5e477-ef42-4c2e-9adc-34d9bf6224d6
  - Presentation Video https://vimeo.com/475446854
- Fiona Smith also gave a presentation at that meeting on hyperspectral error correlations. This was subsequently refined and presented at ITSC in 2020.
  - A recording is here: https://cimss.ssec.wisc.edu/itwg/itsc/itsc23/agendas/orals/ITSC23_Day4_Part2_Recording.mp4 (*16mins and 30s* into the recording)

Observation Error Covariance Estimation methods used by NWP Centres

It had been previously noted that there is an overwhelming predominance of the use of Desroziers relative to Hollingsworth-Loennberg (H-L). The suggestion is that this is the more convenient method within the context of a DA system. For example, it can also give information on background errors and to compute horizontal correlations, so can contribute to selecting thinning distance.

ECMWF, however, do use H-L to initialise the covariances for the first iteration of Desroziers. Other centres appear to initialise the matrices via O-B statistics either for a full or diagonal

matrix. F Smith stated that Hollingsworth-Loennberg is very convenient for full-spectrum analysis because you don't need to assimilate the channels

There is uncertainty on the best initialisation method and this will be important to address for new instrument designs (e.g MTG-IRS). Although Desroziers in theory requires an accurate estimate of B, centres are generally satisfied by impact on forecast skill of Desroziers-derived matrices. V Guidard reported that there are some issues in the atmospheric chemistry context where the horizontal scales in the B-matrix are as in R. F Carmiati reported that in his experiments using Desroziers with CrIS, if you don't have a perfectly defined starting point with B, iterations will begin to drift and it is better to optimise both B and R together. Starting with a very low standard deviation, iterating does not work well. C Harlow has found that a 1D-Var B-matrix can be sufficient to initialise a Desroziers process using 1D-Var rather than 4D-Var.

B Campbell questions whether anyone is using the bias estimate that Desroziers also gives, given that undiagnosed bias can show up as correlation. Nobody seems to have looked at this.

Matrix Regularization

Various methods are applied to the diagnosed covariances from Desroziers to improve performance in the assimilation scheme. For instance, ECMWF use an empirical correction on the variances. This correction was tuned though examining forecast impact for different settings. Inflation is possibly necessary to account for neglected spatial correlations , or maybe because in general scene dependency is neglected. Usually it's a pragmatic choice to improve forecast scores and not enough time to research closely. ECMWF have different inflation for IASI and CrIS - again a pragmatic choice. Documented in papers by N Bormann and R Eresmaa.

JMA have looked at error correlation matrices for AHI, initially they chose to just inflate the diagonal. Subsequently have investigated increasing the smallest eigenvalues. Condition number is maybe not so important for correlation matrices for small number of channels (e.g. AHI).

F Smith’s intercomparison showed that you can end up with similar inflation/conditioning regardless of whether you inflate the diagonal or manipulate eigenvalues. -> Magic number of around 1.75.

Other options for investigation include compression around median eigenvalues.

The general conclusion is that most centres would like to know the “best” way to regularise the matrix, but given they tend to work well with some simple adjustment, it’s not easy to find time to look at these things. There was a general consensus that it would be nice to persuade universities to work on this.

Scene Dependency

A few centres are starting to examine this (ECMWF, NRL, Meteo-France). For instance, at ECMWF there is a study underway to use parameterisations to inflate the variance or change the correlations, depending on the scene. There was some discussion but no conclusion on the best way to interpolate between two matrices between low water vapour and high water vapour

content matrices. D Duncan reported that the LUT approach applied for microwave radiances at ECMWF has had some problems.

ECMWF are also beginning to use SW CrIS channels. For this part of the spectrum the instrument noise is scene dependent.

V Guidard reported not much difference between land and sea scene dependence.

Covariances for All Sky

In all sky assimilation the forecast model errors of cloudy scenes tend to dominate the observation error budget because clouds and precipitation are not predictable on small scales across the assimilation window so their local-scale features are not controllable by the control vector.

Dealing with Spatial correlations

Nearly all centres ignore this issue due to the technical challenges. However there is one operational implementation at the Met Office concerning radar data:

Simonin, D, Waller, JA, Ballard, SP, Dance, SL, Nichols, NK. A pragmatic strategy for implementing spatially correlated observation errors in an operational system: An application to Doppler radial winds. Q J R Meteorol Soc. 2019; 145: 2772– 2790. https://doi.org/10.1002/qj.3592

Covariance Matrices in LAMS

This is an area of interest we did not cover in any great detail but it would be useful to understand. Meteo-France have identified differences in representativeness errors for LAM vs Global. Is there any difference in the most appropriate way to derive covariance matrices in convective scale models?

Physical (bottom-up) approach to modelling error covariances

This approach is not being pursued currently by any centre. Although it is extremely useful to gain insight, some of the terms (errors of representation for instance) are difficult to estimate. The process is therefore quite challenging and time consuming. ECMWF reported that this approach did not appear to be better in trials than using covariances derived from statistical methods.The group concluded that it is useful to gain insight. Perhaps a collaboration with universities would be the best way to carry it forward.

Summary

Little progress has been made since the ECMWF workshop, mostly because the matrices that centres use are good enough in operations, and the important remaining scientific challenges (all-sky; scene dependence) are quite hard. The group would welcome opportunities to work with universities to solve some of these problems, and to look at physically based models and contribute to the understanding of which methods of conditioning are best.

Possible University contacts

- Jason Otkin has a current project on all-sky situation dependent correlated observation error in the IR, not sure it's hyperspectral, but he has some nice ideas for obtaining situation dependence (I would leave it to him to reveal them when he's had a chance to try them).
- Sarah Dance at Reading - she had an Msc student working on sampling error affecting error correlations, though without much progress (and has had previous students looking at various aspects of observation errors including conditioning).
- Peter Jan van Leeuwen and Chih Chi Hu are working on all-sky microwave observation errors (based on the all-sky microwave EnKF dataset Niels and I mentioned). Not correlated error, but a really ambitious attempt at a non-parametric error model (e.g. handling multi-modal errors).
- Dave Tobin (SSEC) - CrIS observation errors
- Larrabee Strow (UMBC) - Obs and RT error work
- Craig Bishop - Any terms that affect DA!