ECMWF/NWP-SAF Workshop on the Treatment of Random and Systematic Errors in Satellite Data Assimilation

Organisers: Niels Bormann, William Bell, Patrick Laloyaux, Karen Clarke

Executive summary

The workshop on the treatment of random and systematic errors in satellite data assimilation was held online 2 - 5 November 2020, and brought together almost 200 experts from NWP centres, space agencies, and academia from around the world. It was organised by ECMWF and EUMETSAT's NWP SAF, continuing the strong tradition of workshops aimed at bringing experts together to exchange ideas and views on topics of high relevance to the use of satellite data in NWP.

Dealing with random and systematic errors in observations and models is at the heart of making optimal use of the wealth of information from satellite data to initialise forecasts or to create reanalyses. Errors and uncertainties arise from many areas, such as in the forecast model, the observations, or the observation operators used to convert model fields to observation equivalents. The challenge is to separate the different errors and to deal with them adequately during the assimilation. One of the aims of the workshop was to connect activities in different communities and to identify where NWP can make use of advances made in other fields in the characterisation of observation-related uncertainties. Reanalysis featured as a strong component, where treating biases in observations and models poses its own challenges.

The workshop featured 23 invited oral and 12 poster presentations, capturing the strong progress in recent years in several areas. Most talks were organised in three sessions, covering the characterisation of uncertainties in observations and models, the correction of observation and model biases, as well as the treatment of random errors. Panel discussions on the treatment of biases and the treatment of random observation errors considered the main challenges and priorities going forward, feeding into lively working group discussions.

The following gives an overview of the three sessions and a summary of the main outcomes of the working group discussions. Recordings of all presentations and panel discussions, and all posters are available from https://events.ecmwf.int/event/170/.

Characterisation of uncertainties in observations and models

Knowledge about observational and model errors and uncertainties informs our treatment of these in the assimilation. On the topic of instrumental biases, talks covered the latest developments in the on-orbit characterisation of the CrIS hyperspectral infrared instruments and the Aeolus Doppler wind lidar, as well as the ongoing activities of the Global Space-Based Inter-calibration System (GSICS, a WMO/CGMS initiative).

The development of metrology-inspired approaches to the characterisation of observational uncertainties featured prominently in two of the talks, and the topic was revisited frequently in the panel discussion on biases and in working group discussions.

Representation error, resulting from a mismatch between the model representation of the state and observations, was explored in two of the talks and promising approaches were described involving high-resolution model runs and observations. A talk on the characterisation of uncertainties in historic observations illustrated the value of the long-term perspectives brought by reanalyses.

Correction of observational and model biases in data assimilation

Correcting observational biases is essential for the successful assimilation of many satellite observations. Adaptive bias correction methods are now commonly used within the assimilation system. They are based on the specification of observation bias models that remove the systematic error when the observation departures are computed. Several talks discussed the best way to build these observation bias models to avoid absorbing model error in this correction (e.g. Constrained VarBC, anchor-only reference state, better choice of predictors).

Separating observation and model biases continues to be challenging. Several presentations discussed the concept of scale separation where model error contains identifiable large-scale structures, which opens a new perspective in the quest to attribute the correct source of biases. Machine learning methods trained on analysis increments or anchor observations also showed some potential to learn the correct structure of model biases.

All the talks illustrated the importance of anchor observations and more specifically GNSS RO to estimate the different types of biases. In this context, reanalysis is facing a challenge, with highly variable observational coverage that becomes sparse going further back in time.

Treatment of random observation errors in data assimilation

The random error characteristics assigned to observations in the assimilation play a key role in determining the weighting of observations in the assimilation. Aside from measurement errors, many satellite observations are also strongly affected by representation error, that is, errors that occur when model fields are mapped into observation equivalents.

Current observation error modelling mostly relies on diagnostics based on differences between observations and model equivalents from short-range forecasts or analyses, and several speakers highlighted the potential, but also pitfalls of these approaches. Situation-dependent observation error models have been in use for some time, for instance for Atmospheric Motion Vectors. Here, simple parametrized models are employed that capture main variations in error contributions, for instance, from larger errors in the forward modelling.

Many centres now account for inter-channel error correlations, particularly for hyper-spectral IR instruments, and a dedicated presentation reviewed commonalities and differences. Concepts for modelling situation-dependent inter-channel error correlations are being developed, for instance in relation to cloud-effects in all-sky assimilation. A better understanding of the effect and treatment of spatially correlated observation errors is emerging, with first indications of benefits from handling these, either explicitly or through pragmatic approaches. However, algorithmic challenges remain, especially for variational data assimilation systems.

Main outcomes from working group discussions

Parallel working groups were held, with two working groups each considering the treatment of systematic errors and the treatment of random observation errors, respectively. They reviewed the state-of-the-art presented at the workshop and discussed possibilities and priorities for future developments. While the parallel working groups on the same topic naturally developed different areas of emphasis, the main outcomes were largely consistent. The following gives a summary of these main outcomes; full reports from each working group are given in the appendix.

An overarching recommendation from all four working groups is the call for more work regarding metrological/physical understanding of systematic as well as random observation-related errors, as it is seen as fundamental in informing their treatment in data assimilation. This covers instrument-related errors, as well as representation errors, arising, for instance from observation operators (e.g., radiative transfer) or spatial representation errors (e.g., linked to turbulent scales). For radiative transfer models, it is recommended to investigate separately errors in fundamental line-by-line models, as well as errors due to instrument design uncertainties (e.g., response functions). More effort is also needed to utilize available information from metrological/physical error analysis in the treatment of random and systematic errors in data assimilation.

For the treatment of systematic errors, the working groups further identified the following main items (in no particular order):

1. There is a need for better reference observations or fuller error characterization to anchor bias corrections and to better identify model biases, particularly for tropospheric humidity, high-altitude (mesosphere) and higher-depth (deep ocean). Hyperspectral IR could be considered as anchor measurement, potentially in conjunction with Constrained VarBC and better uncertainty estimates. Traceable uncertainties for GNSSRO and hyperspectral IR are required.
2. The workshop encourages community efforts to better analyse and address observation-related biases, for instance, through inter-comparison of bias corrections from different centres (incl. bias models used), open sharing of databases with feedback information, and better dialogue between NWP centres, GSICS and instrument designers. Continued effort should be made to identify biases at source.
3. There is a need for NWP centres to revisit bias models used in VarBC to reflect advances in reducing systematic errors. Further work is also required to investigate correlations between estimates of observation bias and estimates of model bias in observation space.
4. More work is recommended to characterize uncertainties in climate reanalyses. Observation denials and model parameter perturbations combined with ensemble approaches are suggested as promising avenues to characterize specific aspects. It is recognized that producing unbiased analyses is a higher priority for reanalyses compared to NWP, hence calling for dedicated developments in certain areas (e.g., more extensive use of CVarBC).

For the treatment of random observation errors, the working groups identified the following aspects as key priorities (in no particular order):

1. There is a need to better understand the diagnostic uncertainty estimation tools and the estimates that they produce, including understanding the influence of background and model error on diagnosed observation errors. Cross-comparison of results from different tools is recommended, as well as comparison to metrological/physical estimates.
2. The groups recommend developing further the treatment of situation-dependence of observation errors, including the treatment of situation-dependent error correlations where appropriate. Results from departure-based diagnostics may have to be treated with extra care in this case, due to increased sampling error when splitting the error estimates into different situations.
3. The groups recommend increased efforts targeted at overcoming the technical challenges that currently limit the use of horizontal error correlations. This is seen as a particular priority for convective-scale systems to better assimilate small-scale features.
4. More work is required regarding automated or online estimation of observation errors. This is considered particularly important when dealing with many new satellite instruments simultaneously, such as future constellations of small satellites.

The working groups concluded that building on the strong progress in recent years, there is ample scope for further improvements in the treatment of random and systematic errors in satellite data assimilation. In addition, a growing and increasingly diverse observing system as well as developments in assimilation methods pose further challenges for the treatment of systematic and random errors. Hyperspectral sounding data with unprecedented temporal resolution will be available from geostationary satellites, observations from some smaller satellites may have less well-characterised uncertainties, and an increasing number of observations will be driving multiple Earth system components with different model error and levels of maturity. All of these aspects will require continued development of the methods and capabilities used, including for NWP-independent characterisation of all error sources.

Appendix: Working group reports

Working group 1: Treatment of biases I

Co-chairs: Paul Poli (EUMETSAT), Patrick Laloyaux (ECMWF)

Participants: Mazlar Bani Shahabadi, Bill Bell, Massimo Bonavita, Mark Buehner, Xavier Calbet, Paola Corrales, Dick Dee, Amal El Akkraoui, Devon Francis, Sergey Frolov, Timo Hanschmann, Viju John, Christina Köpken-Watts, Christopher Merchant, Rob Roebeling, Roger Randriamampianina, Dinand Schepers, Qiwen Sun, David Tobin, James While.

This session took into account questions raised during the previous days, on the topic of systematic errors. Feeling a general appetite for tackling this issue at the source(s), the working group discussion addressed five main topics: (1) diagnosing biases, (2) bias sources, (3) reference data, (4) bias correction methods, and (5) NWP *vs* climate applications of data assimilation.

1. Diagnosing biases

Biases matter when they degrade the products

Participants noted that agencies diagnose biases in any new component of the observing system before declaring operational assimilation. The goal is to verify that existing methods detect biases in observation minus background (or analysis) departures, and correct residual signals, if needed. Given their priorities, teams handling the assimilation of several instruments only grant further attention to the topic of bias if necessary, e.g. lingering bias, NWP forecast score degradation. Some reanalysis biases were also addressed in NWP. This explains the present organization of the diagnostic approaches, guided by data assimilation departures, at NWP centers.

Limits of this approach, and advancing towards 'error inventories'

The first goal of the data assimilation process, i.e., improve a prior background state to deliver an analysis used for the subsequent forecast, justifies this situation. However, the imperative improvement of the forecast quality may prevent investigations, from the data assimilation side, to quantify (even residual) biases in the analysis. Consequently, when an unexplained bias does show up in one component of the observing system, the approach to address the diagnosed bias is very much *ad-hoc* and guided by pre-existing communication channels between the data producers and the NWP centers. However, these channels vary greatly. Participants recognized that existing international frameworks for such discussions are adequate (e.g., ITWG for the atmospheric sounders), should one use them for this purpose also. Indeed, in these fora, the NWP centers and data providers may find more occasions to more systematically discuss together what they could explain in diagnosed/presented biases, and **to gradually develop, together, error inventories, starting with instrumental uncertainties**.

Bias diagnostics resources online, and towards greater data exchange

Presentations showed a significant amount of pre-generated diagnostics on the web, featuring systematic differences between observations and model backgrounds or analyses or other observations. Such resources are located both on the data producers' and data users' sides, including also NWP-SAF and GSICS, and cover several instruments. Each set of diagnostics comes with its own specificities in terms of instruments, details, timeliness, and application. While efforts by multiple agencies enable to track possible failures in the product dissemination chain, these varied sets of plots often raise many questions, the differences between them having generally to do with different data populations. Also, these plots generally do not offer much perspective in terms of physical context. A next step to help answer those questions would be to allow users to drill down, from the web plots, into the data that populated them. This would then let the users of these diagnostics further refine/recompute statistics by criteria of their choice or other information at their disposal (e.g., about instrument status, or about the atmospheric/oceanic/terrestrial conditions at the scene). A possible concrete development towards this direction would be **to open up the access to databases of observation diagnostics statistics and underlying data**. This would start a novel data exchange method between data producers and data assimilation users, where each party would include the contextual information at its disposal. Such infrastructure could also help accelerate the evaluation of future mission (or reprocessed) data.

2. Bias sources

Instrument systematic errors

Citing examples of successful bias investigations, the group noted the preponderant role of 'physical intuition'. However, this 'top-down' approach leaves the outcome to expertise availability and chance, and scales poorly with an increase in the number of instruments or platforms (e.g. future missions) that need to be evaluated before operational assimilation. **A 'bottom-up' approach would estimate the structures and magnitudes of maximum bias one may expect for the instrumental error, using instrument bias simulations, similarly to the FIDUCEO project** (measurement equation and table of effects).

This appears easier to **initiate with new sensors** (and may be pursued with, e.g., IASI-NG). For new instruments, simulators are generally developed from the onset, in collaboration with the instrument designers, taking into account all the instrument components and studies results that may not always be openly communicated otherwise for practical reasons. **For historical climate sensors (e.g., MSU), case studies would be relevant**, revisiting the measurement equation and drafting tables of effects as part of the reprocessing.

Representation systematic errors

An important error source in the case of IR or MW sounders is the departure of the spectral response function or central frequency from nominal specifications. **Participants expressed interest in knowing the accuracy specified for these parameters**. A further need is the **ability to propagate this uncertainty into the observation operator** (RTTOV). Also, such **metadata information flow should be organized and kept consistent with the WMO OSCAR/Space** database.

Similarly, participants expressed interest in estimating the structures and magnitudes of systematic errors potentially caused by horizontal and vertical inhomogeneity. This calls for **simulation studies, using two types of input, from high-resolution observed data (e.g., observation campaigns, high-resolution instruments…), and from very high resolution model data (at finer scale than regional NWP)**. Another source of systematic error resides in the spectral domain. Simulation studies could also help estimate the magnitudes of errors due to spectroscopy and fast models, by comparing with line-by-line radiative transfer models, and communicate results to data assimilation users. Model and forcing systematic errors were discussed with climate applications in mind (see last section).

Consistency check for all systematic error sources

Verifying that magnitudes and structures of the various systematic error sources are adequately diagnosed and corrected may be pursued in data assimilation system experiments along the lines of **Observing System Simulation Experiments (OSSEs)**. This allows controlling all error ingredients, and verifying outcomes against expectations.

3. Reference data

Reference data sources

Several types of observation data are already available and serve as references in most data assimilation systems. GNSS-RO is used as an anchor in the stratosphere in NWP and reanalysis. Radiosonde observations are used to anchor the troposphere and the lower stratosphere, noting however that these observations are sometimes subject to a bias correction to remove small biases (e.g., radiation) in the observation pre-processing. Argo observations serve as anchors for the ocean models, using a fleet of robotic instruments that drift with the ocean currents and move up and down between the surface and a mid-water level. Finally, climate forcings used in Earth system modelling have generally been derived from observational data and are used to control model drifts, typically without bias correction, i.e., model forcings are effectively treated as reference data.

Improving the use of existing reference observations

Participants questioned whether or not GRUAN data should be bias-corrected at all in NWP. Similarly, **trials to use hyperspectral IR as uncorrected references should be pursued, at least for the temperature channels**: the residual bias could be handled using a constrained variational bias correction approach. Humidity channels are more challenging to treat as references because of spectroscopy and representation errors, and suspected multivariate model biases. In addition, **data collected by observation campaigns (e.g., research experiments, satellite cal/val) are probably insufficiently used at the moment**. Such data, often from research-grade, well-calibrated instruments, could provide useful resources for offline research to understand NWP bias sources.

Need for more error-characterized observations

Noting the present sparsity and limited timeliness of in situ reference data, the group acknowledged the need for more reference data in routine mode. **Anchoring NWP systems requires reference data for (especially) tropospheric humidity, high-altitude (mesosphere) and higher-depth (deep ocean).** It was also noted that any component of the observing system treated as reference should have its errors traceable and characterised. Indeed, **some existing observing systems, effectively treated as references, are still missing rigorous traceable uncertainty estimates (e.g. GNSS-RO)**.

4. Bias correction methods

General framework: variational bias correction (VarBC)

Data assimilation algorithms generally assume that observations present random zero-mean errors, which implies that all observations are considered as references, unless the formalism makes provisions for an explicit representation of biases. Different approaches have been developed for this purpose, to estimate and remove observations biases (from fully-controlled bias to completely unknown bias). Online VarBC is the current mainstream approach, but other offline methods are also promising. For example, ingesting only reference datasets can provide a bias-free trajectory for the observation bias model of the main system assimilating all the other observations.

Going further with existing VarBC

Several possibilities were noted to improve upon current VarBC. A renewed collaboration between the different NWP and research centres with extensive experience with VarBC would probably be valuable. For a start, **VarBC corrections and model predictor bias definitions could be exchanged between centers and compared for the different types of instruments, including with other independent estimates (e.g. GSICS)**. Second, **revisiting the choices of predictors used in the present implementations of VarBC is long overdue**. A dedicated study would ideally be internationally coordinated, between the various data assimilation centers.

New possible approaches based on VarBC

The background error term in VarBC requires more attention, as it defines how fast the bias correction reacts to changes in mean departures (known causes may be instrumental or model-based). **An ensemble approach (ensVARBC) could be developed to characterize the uncertainties of the correction**. The constrained VarBC (cVarBC) method seems particularly suited to handle reference data, as it allows controlling the magnitude of residual bias after assimilation. To make progress, **cVarBC would benefit from uncertainty estimates (ideally) produced by the data providers, and constrain accordingly the amplitude of the bias correction**.

Disentangling biases: on the importance of assumptions for diagnosing sources

Separating observation and model biases continues to be challenging. Several presentations discussed the concept of scale separation, where model error contains identifiable large-scale structures, which opens a new perspective in the quest to attribute the correct source of biases. This concept supported the recent improvement achieved by weak-constraint 4D-Var. However, this is not the only possible algorithm. For example, **a weak-constraint 3D-Var may be able to estimate the model bias, even if adjoint/tangent codes are not available**. The model bias can also be estimated from the increments produced by the anchor-only assimilation system and the correction be applied to the model equations in a similar way as weak-constraint 4D-Var. Enabling further work to disentangle biases calls for inventories of systematic error effects, with expected structures and magnitudes, to develop algorithms able to infer the bias source in a simulated environment.

5. NWP vs climate applications of data assimilation

Biases matter more to reanalysis

Issues posed by biases are particularly acute for reanalysis because they can compromise the usability of reanalysis products for climate studies. However, participants felt that NWP also would gain from pursuing understanding of the (analysis) biases in present-day NWP. To achieve this, **a reanalysis of the present times, based on the best observations, would provide a benchmark, aiming to be uncertainty-characterized**. From the benchmark baseline, subsequent reanalyses with stochastic perturbations of key model parameters, and with varying amounts/types of observations would further help to characterise biases, as follows.

Understanding the impact of model biases on reanalysis trends

Improvement of climate signal representation in reanalysis may be obtained by **following common protocols similar to what is done in the CMIP community**. Participants also noted that **model-only runs are essential** in the basket of reanalysis products, to guide producers and users in the origin of possible biases. In addition, **low-resolution reanalyses with varying model parametrizations** would help explore, gradually, the space of model biases and their impact in reanalysis climate timeseries.

Understanding the impact of changing observations on the reanalysis trends

Participants noted a tension between two reanalysis objectives: delivering 'unbiased' and 'best' solutions for the past, and delivering continuous initial conditions for reforecasts (e.g., for seasonal forecasting or extreme forecast indices). The intrinsic methodology of reanalysis makes it difficult to realize both at the same time. One trade-off would be to **deny some key observation datasets** (or components of) and retain them for validation, in additional low-resolution reanalyses. This approach is not systematically explored enough today. Similar to the previous point, this would help gradually explore the space of observation biases and their impact in reanalysis climate timeseries.

Learning from reanalyses for future NWP improvements

Participants maintained that reanalysis remains extremely valuable for NWP in general and also specifically regarding assessment of biases. As an example, ERA5 led to a better understanding of the stratospheric model biases. Reanalyses can also help to specify new observing systems, addressing identified/past flaws. For example, regardless of other developments, **in-orbit calibration missions (such as TRUTHS or CLARREO) remain needed**. Also, the **reanalysis community may help review the GCOS/WMO requirements** and ensure that the observation needs of future reanalyses are covered. Finally, it is noted that **the observation feedback archives from reanalyses are still mostly unexplored resources** that contain much of information on systematic differences between observation and model, including (for some observations) bias correction timeseries.

Working Group 2: Treatment of biases II

Co-chairs: Stuart Newman (Met Office), Sean Healy (ECMWF)

Participants: Prashant Kumar, Kozo Okamoto, Tim Hewison, Reima Eresmaa, Indira Rani, Fabien Carminati, Thomas Hall, Amos Lawless, Marco Matricardi, Pieter Houtekamer, Ruth Taylor, Robin Faulwetter, Kamal, Sanita, Peter Steinle, Rob Kursinski, Larrabee Strow

The group addressed the following talking points.

1. How well can we separate biases arising from (1) biases in the observations and/or forward models and (2) biases in the NWP model

1.1. Are present anchor observations sufficient?

There was a clear consensus that having a greater number of anchor observations would help to constrain model biases.

The group considered which observations are routinely used as anchors (principally GNSS-RO, AMSU-A channel 14, conventional observations) and which might be suitable candidates in the future. Key requirements for anchor observations are robust uncertainty characterisation, long-term stability and sensitivity to the geophysical parameter(s) in question.

Channels from hyperspectral infrared satellite instruments are considered to be promising candidates to act as anchors, potentially for temperature and humidity in the troposphere, where GNSS-RO has less influence. Progress in developing radiative transfer models over the last 10-15 years (e.g. new spectroscopy), along with measures such as using a realistic CO_2 profile, mean that biases can be reduced to “acceptable levels”. By acceptable levels, we mean biases that are small when compared with the assumed observation error statistics used in the assimilation (e.g., <10%), rather than requiring the bias to be zero.

Recommendation: Explore the use of hyperspectral infrared satellite data as anchor measurements.

It was noted also that we can now intercalibrate geostationary radiances with high quality references such as IASI and CrIS, and these could be explored as anchor measurements depending on the application (requiring targets for uncertainty and stability).

The group agreed that it is important to fully characterise biases in the line-by-line spectroscopic models underpinning radiative transfer models such as RTTOV. This could be supported using high quality in situ data (GRUAN radiosondes for example) for forward modelling in comparison with recognised well calibrated satellite observations.

Recommendation: Investigate and characterise biases in fundamental line-by-line models separately from the forward model

1.2. What is the role of further constraints, such as bias models in VarBC, model error covariance statistics, implicit constraints from parameter estimation? What can we do to specify these better?

It was noted that in applying bias models we make assumptions about the forms of the observation bias and model bias. In theory, the model error term in weak constraint 4D-Var and VarBC should be correcting different error sources, and not be correlated in observation space, but this should be tested.

Recommendation: Investigate correlations between observation biases and model error terms in observation space.

The group agreed that some aspects of the bias correction framework in NWP date back many years, and it may be timely to revise the form(s) of bias predictors used. It was noted that airmass predictors were devised at a time when forward model and instrument errors were generally larger than we find today.

Recommendation: Reassess the role of airmass and other predictors in modern VarBC applications.

There was also interest in the group in exploring the use of robust instrument calibration uncertainties, such as those determined by the Fiduceo project, in schemes such as constrained VarBC.

1.3. How successful are current methods in an all-sky framework?

No clear recommendations in this area.

1.4. Further challenges in coupled Earth System Assimilation systems?

The group was concerned that we need to understand better how to bias correct observations that have sensitivity to both the atmosphere and ocean in coupled models. The example of scatterometer winds was noted, which has a sensitivity to both low levels winds and the ocean currents. There is a concern biases from one component in the earth system may cause, or reinforce, biases in another.

2. Estimation of model bias (either during the assimilation or through increment analyses etc.)

2.1. What techniques look most promising? E.g., model parameter estimation or model tendency correction? Are they mutually exclusive?

A strong interaction between model developers and data assimilation/observation specialists in NWP would be beneficial to future efforts to tackle model bias. Model developers can help inform priorities for investigation. Parallel efforts should continue to 1) improve models and reduce the model errors, and 2) estimate the model errors within the DA system.

2.2. What can we learn from the corrections for model development?

Some parameters could be improved or constrained using data assimilation techniques, but it is recognised that not all are well constrained by the available observations. Members of the working group reported some success, estimating parameters from DA systems, but also reported difficulty subsequently connecting these results to new model developments. It was also recognised that some model biases would be easier to correct with DA techniques than others. Biases associated with convection were described as “horrendous”, but others associated with biases in land datasets might be easier to deal with.

2.3. Should we apply the model bias corrections derived in the assimilation during the subsequent forecast?

Recommendation: Test weak constraint and parameter estimates in forecast mode.

2.4. What are the tools available to diagnose model biases over different timescales (e.g. assimilation window, medium-range, seasonal)? How can we estimate higher-order statistics of the model error (e.g. covariance)?

GNSS-RO observations are seen as the primary source of high quality monthly mean biases and higher order statistics in the stratosphere.

Stochastic parameter perturbation schemes are designed to give a realistic spread in ensembles, but they should provide useful information about the growth of the “real” model error. This information should be useful for estimating model error covariance matrices.

3. Estimation of observation/observation operator bias

3.1. What can we learn from bias corrections about addressing biases at source?

The group supported the investigation of bias statistics from NWP or GSICS in support of root cause analysis allowing biases to be corrected at source. The recent studies with Aeolus were seen as a good example of careful analysis, enabling the bias root causes at the instrument level to be identified and understood.

While there are inevitable forward model biases caused by errors in the underlying spectroscopy, these biases can be minimized by keeping up to date with the latest state-of-the-art radiative transfer developments.

It was noted that there is a situation dependence in some forward model biases, for example in the use of atlases to specify surface characteristics.

3.2. What independent estimates of observation bias do we have? Are we making full use of them?

Examples of independent sources of observation bias information include pre-launch error budgets, intercalibration during the mission lifetime (e.g. GSICS) and post-launch calibration uncertainty estimation (e.g. Fiduceo). As noted in Section 1.2, exploring the use of calibration error budget information, to provide realistic ranges of biases, in constrained variational bias correction schemes was seen by the group as a way of making better use of the available information.

3.3. Changes in observation system coverage in climate reanalyses affect trend estimates. How can we mitigate this?

The group saw promise in assessing reanalysis performance with differently constituted observing systems, during a well observed, recent period. One way of achieving this, raised during the workshop, is to withhold a subset of very high quality observations for the purposes of validation, and to test the sensitivity of the reanalysis state to removing these observations.

3.4. Do we need different bias constraints for reanalysis?

It was noted that short- and medium-range forecasts and reanalysis have distinct goals (best forecast versus best analysis) and that these may not always be optimised by the same bias constraints. Exploring the use of constrained VarBC schemes using instrument uncertainty budgets may be a higher priority for reanalysis compared to operational NWP. Another potential constraint for reanalysis is the estimated instrument stability derived from satellite intercalibration studies. Giving more weight to the anchor measurements in reanalysis could also be tested.

4. Is there anything more that could be done to accelerate progress (towards a bias-free world!) - through ways of exchanging information, coordination, prioritisation…?

The NWP community gathers valuable information about instrument bias characteristics through targeted studies of first guess departure statistics post-launch and through routine monitoring. There is the potential for greater data exchange between NWP centres and WMO/CGMS bodies such as GSICS on error budgets, instrument bias characteristics and NWP departure statistics. Information on outages, sudden data quality changes, drifts over time and blacklisting can flow both ways.

Recommendation: establish greater dialogue between NWP centres and GSICS leading to a systematic data exchange system for biases and alerts.

Spectral response functions (SRFs) should be made routinely available, possibly on the WMO OSCAR pages. There is also a requirement to make better use of uncertainties in SRFs where these are characterised. These uncertainties need to be mapped into the observation error budget, e.g. via a radiative transfer model.

Assessments of instrument biases independent of NWP are also useful. The well calibrated observations at the GRUAN sites could also be used to assess observation biases.

5. Are there future challenges (e.g., the evolving satellite observing system, the move to Earth System Models and DA, …) that present particular challenges?

As we enter the era of small satellites with a short lifetime, there is a need for rapid appraisal of instrument bias characteristics so that sensors can be used in operations quickly post-launch.

The group expressed concern about future infrared hyperspectral instruments in the global observing system which rely on shortwave infrared only channels. These instruments exhibit complicated bias characteristics due to issues such as non-Local Thermodynamic Equilibrium (non-LTE) and the requirement to model the contribution of solar radiation.

Working group 3: Treatment of observation errors I

Co-chairs: Nancy Nichols (University of Reading) and Pete Weston (ECMWF)

Participants: Alan Geer, Guannan Hu, Kristen Bathmann, David Duncan, Stefano Migliorini, Jemima Tabaert, Pierre Gauthier, Alison Fowler, Jonathan Mittaz, Kelvy Cardoso, Shuang Xi, Koji Terasaki, Fiona Smith, Bill Campbell, Paromita Chakraborty, Prateek Dongre, Kirsti Salonen, Cristina Lupu, Federico Cossu Tony McNally, Saleh Abdalla

The objective of Working Group WG3 was to consider the treatment of random observation errors in data assimilation. The working group focussed on four topics raised by ECMWF. Lively discussions were held on all the topics and a variety of issues were raised for future study.

1. Estimating Observation Errors

1.1. What tools do we have to estimate observation errors and how well do they cover our needs?

Estimating the observation error covariances is difficult due to the complicated relationship between the observation and background errors affecting the assimilation. The most widely used error estimation techniques are the departure-based methods such as the Desroziers and Hollingsworth-Lonnberg methods. Another approach subtracts some independent estimate of the background error (e.g. from EDA ensemble statistics) from the innovation covariance. There is also a triple collocation method, but the members of the working group had little experience of using this. Finally, there are the metrological approaches, where an error estimate is built up from individual estimates of the constituent parts of the observation error (instrument noise, forward model error, pre-processing error, representativeness error). Discussions of the pros and cons of these techniques were held at length and are summarized as follows.

1.2. Adjustments to diagnostic observation error estimates: why do we need to make them?

There are many different adjustments that are necessary in order to use diagnosed observation error covariance matrices in assimilation systems. Some of these are required mathematically for the diagnosed matrices to be converted into valid covariance matrices and some are more pragmatic approaches to enable the assimilation algorithms to converge satisfactorily.

Firstly, the diagnosed matrices often must be corrected to ensure that they are valid symmetric and positive-semidefinite covariance matrices; this can be done by straightforward mathematical processes, but these introduce additional errors that should be considered. The next set of adjustments that are often needed involve modifying the eigenvalues of the diagnosed covariance matrix to improve its conditioning; this is referred to as reconditioning or shrinkage. This procedure has three main motivations:

a. Mathematical: Making the covariance matrix easier to invert accurately, with less computational work.
b. Pragmatic: Improving the conditioning of the Hessian, which enables the optimization procedure to converge more quickly in the assimilation system.
c. Physical: Reweighting the departures. Inter-channel first-guess departures that map onto eigenvectors with small eigenvalues are given large weights in the assimilation. In practice these eigenvectors often have highly oscillatory structures that may lead to unphysical analysis increments or features in the analysis that the forecast cannot correctly represent (e.g. gravity waves) and thus degrade the analysis and forecasts. These effects can be reduced by increasing the smallest eigenvalues and thus giving observations which map onto these structures less weight.

There are several different methods for performing the reconditioning but the main two are: ridge regression, which involves adding a fixed constant to the eigenvalues; and eigenvalue floor, where all eigenvalues below a certain threshold are increased to that threshold value. Different centres have used different methods depending on their motivations and both seem to perform similarly. One concern is that the blocks of the covariance matrix representing different types of channels are quite different, so the question arises: could benefits be gained by making different adjustments depending on the channel? A point was raised in the discussion about which matrix we should perform the reconditioning on, with a preference for applying it to correlation matrices rather than covariance matrices so that the standard deviations are preserved.

What do the adjustments tell us?

The effects of the adjustments are not fully understood and more should be done to try to understand the modifications that are made. Covariance matrices represent physical signals. In reconditioning the diagnosed covariances, eigenvectors with small eigenvalues that represent some features are then given larger weights. These may be instrument features rather than geophysical features, as has been found in work on principal components by Tim Hultberg. The eigenvectors of the PC compression have been shown to map onto a mix of physical features and instrument characteristics.

The level of asymmetry in the diagnostics could suggest other problems in the assimilation system (e.g. B, residual biases, model error feeding into observation error). Better methods for separating random and systematic errors are needed. There was a suggestion of using a neural network approach to avoid the problem of estimating B for use in retrievals. Also, better estimates of B should lead to improved estimates of R due to their complicated relationship in the diagnostics.

A few examples of this were given showing the sensitivity of the diagnostics to other problems within the assimilation system. In one example, a very highly correlated group of water vapour channels was found. The eigenvectors representing combinations of these channels appeared to map onto gravity waves and had tiny eigenvalues. A pragmatic approach was to pick one channel and discard the others, which had little influence on results but led to better convergence. There are also examples of rooting out channels with large residual biases and refining surface type quality control based on results from the diagnostics.

In general, a healthy scepticism of the Desroziers diagnostics and other departure-based diagnostics was expressed. One example found negative variances when subtracting an estimate of background error from the innovation covariance, suggesting problems with the background error estimate. For CrIS the standard deviation of O-B is less than 0.1K, which includes observation error and background error. Off-diagonals are of order ~0.01K so it was argued that our current knowledge of B is not good enough to estimate these numbers accurately. There are many degrees of freedom in 100x100 matrices that we are attempting to quantify. Could a physical, parametrised approach similar to that done in all-sky assimilation, with an error model based on cloud predictors, be a better approach?

Despite the problems noted above, all NWP centres have found significant improvements to forecasts from using inter-channel error correlations. Results suggest that using an estimate of the correlations is better than nothing and definitely better than the uncorrelated and inflated observation errors used previously. However, if we address the concerns above, then there is potential for further improvements from the optimal specification of observation errors in the future.

1.3. Uncertainty characterisation beyond departure-based diagnostics (e.g., instrument characterisation, metrological approaches, etc) - how is that feeding into observation error specification?

There was a lot of support for increased usage of metrological approaches where error estimates are built using a “bottom up” approach. **The group recommended more work in this area.** For instrument errors we can consider the raw measurements (voltages) and build up uncertainties from these. For example, an error in the temperature of a calibration target leads to correlated error for all channels that use that calibration target. Estimates of inter-channel covariances of instrument noise are available for some instruments (IASI, CrIS, ATMS), but can these be provided by more instrument providers? The radiative transfer model (RTM) community do comparisons to quantify errors that could feed into metrological models. How these error estimates are communicated to our community could be improved. One major difficulty with metrological approaches is how to estimate representativeness error; for this it is still necessary to resort to departure-based methods. One suggestion was to use large eddy simulations of turbulence as a measure for sub-grid variability and to use that to estimate the representativeness error. Could we use parametrisations in the model to deal with this? Using this approach could lead to a more flow dependent error estimate. Pre-processing and quality control errors are also difficult to quantify accurately. There was a suggestion to split up the specification of observation errors into constituent parts. Also, there was a suggestion to use departure-based methods, subtract known contributions from metrological estimates and study the remainder more carefully.

2. Accounting for Observation Errors

2.1. Status of accounting for spatial/temporal error correlations

There has been considerable attention recently on methods for treating inter-channel error correlations, but thinning, variance inflation, and superobbing are currently the most common approaches to spatially correlated observations. Allowing for spatial error correlations would enable significantly more data to be assimilated, but there are technical, computational challenges to overcome. This issue will become more important for higher resolution and convective scale models. **The group concluded that more work should be done on practical approaches for the implementation of spatial error correlations.**

For temporal error correlations if the innovation vector is kept ordered in time it should allow for a block diagonal R matrix. For satellite observations there is a link between the spatial and temporal error correlations. There is the potential for deriving spatial error correlations from the instrument, e.g. a single calibration for geostationary satellites is used for an entire global scan. The usual correlation-distance model may not hold for more complex systems such as strongly coupled Earth system models, particularly at the boundaries between different components. A way forward could be to use a parameterized approximation that is easily invertible e.g. Toeplitz/circulant matrices. We need to be careful not to confuse spatial or temporal correlations with biases, e.g. cross-scan variability, which is mostly dealt with by bias correction.

2.2. When and why do correlations matter?

Correlations matter when we want the assimilation system to respond to differences between observations. We've seen examples of this with inter-channel error correlations: for example, when there is a broad innovation of the same sign in all channels, this is more likely to be an observation error, so this observation is down-weighted; conversely when the innovation switches sign between correlated channels this is more likely to be a background error so this observation is up-weighted. Similarly, spatial error correlations will tell the assimilation what weight to give features in the spatial differences between observations; this links to Joel Bedard's difference observations approach. The correlations need careful treatment in relation to the spatial and temporal filtering. In the early 4D-Var approach, surface pressure tendencies were assimilated without taking account of temporal correlations which resulted in the model blowing up. It will be more important to take spatial and temporal error correlations into account when we go to higher resolution models or in convective scale models. This will allow us to better use higher resolution and more frequent observations to constrain these models.

Alternative approaches

We generally expect that positive spatial correlations will lead to down-weighting observations; therefore, the simpler approach is to superob or thin the data. Superobbing leads to reduced instrument noise and representation error, but increased cloud contamination. Cloud cleared radiances approach could help but leads to more correlated errors due to correlations between observations and background, which violates the DA and diagnostic assumptions. There is also the approach of assimilating difference observations.

2.3. Correlation between observation errors and background errors

One of the fundamental assumptions in modern data assimilation methods is that the observation errors and background errors are uncorrelated. However, there are possible mechanisms for introducing such correlations. For example, in the ensemble of data assimilations (EDA), perturbed observations are used to maintain the spread between ensemble members. The outputs from this system contribute to the hybrid background error used in the assimilation. In addition, background fields are used extensively in the quality control of observations, which could lead to a circular dependence and correlation between the background and observation errors. In the past, satellite retrievals based on a priori background information led to strong cross-correlations. Avoiding this issue was a contributing factor to the development of direct radiance assimilation in variational methods. **The issue of correlations between background and observation errors should be investigated further.**

3. Situation-dependence of observation errors

3.1. Situation dependence

This topic generated an avid discussion. **There was a strong view that more centres should look at expanding their use of situation dependent errors,** particularly for aspects such as different surface types or latitude bands.

NCEP use different error correlations over land and over sea with generally larger errors, stronger correlations, and some anti-correlations over land due to surface emissivity uncertainty, skin temperature biases and interactions with the cloud detection. Other centres use a fixed correlation matrix but use quality control to filter out observations with larger errors e.g. surface-sensitive channels over land. Using situation-dependent errors should be an enhancement on top of this approach. All-sky has benefited from situation dependent uncorrelated errors with larger errors of representation in cloudy areas. There is ongoing work to combine situation dependence and correlated errors. The eigenvalue approach from Alan Geer for all-sky IR where only the leading eigenvalues of the matrix (sensitive to cloud) are inflated is a promising approach, but the technique won't be operational soon due to other parallel developments. A parametric approach is being pursued in clear-sky IR assimilation, where stronger correlations are diagnosed in channels sensitive just above the detected cloud.

Situation dependence based on meteorological conditions is harder to implement and will require more resources. There was a suggestion that we could measure the variance or uncertainties (confidence intervals) in the error estimates. Larger variances in global estimates could suggest the need for more situation dependence.

3.2. Level of sophistication - what is needed

More sophisticated error models will potentially lead to more maintenance of the current operational system. For example, how often would the observation error covariance matrix need to be re-calculated based on significant model changes, calibration changes to the observations or other large observing system changes. The background error covariance matrix will be updated via changes to the EDA ensemble system but there is no routine update to the climatological part; should there be?

Also, the question arises as to whether the diagnostics are accurate enough to estimate scene dependent correlation structures when the samples used will be smaller than for globally averaged estimates. Finally, there is a danger of aliasing model/background errors into observation error estimates, which could be made worse when looking at smaller samples based on different parameters and predictors.

Convective-scale models

Different errors are used in limited area models from global models at Meteo France, which is justified by different representativeness errors resulting from the different model resolutions. Also, different model tops can affect the channel selection and error correlations. Different representation errors can arise when model resolution is coarser than observation resolution and vice versa, for example, when the footprint of the observation is larger than the model grid size. Sub-grid variability is important in all-sky assimilation in clouds. Displacement errors become more important at higher resolution; should they be part of observation errors or background errors? The experience of the group was that it is hard to focus developments on convective scale models because experimental results are currently less reliable.

Potential to use machine learning

Could we use a machine learning (ML) approach with different features or potential predictors to work out which aspects of situation dependence are important? There is the potential for model error contamination in this kind of technique. Could ML be used to remove noise or the influence of background errors on estimates? ML has already been used in bias correction techniques - could it be applied to observation errors?

In future we may need to assimilate principal components (PCs) or reconstructed radiances (RRs) for hyperspectral instruments such as MTG-IRS and IASI-NG. The observation error covariance matrices are significantly different for PCs or RRs compared to raw radiances with stronger correlations, non-local response and better conditioning.

Another option is to assimilate transformed retrievals but here the 1D-Var used to produce the retrieval needs a good estimate of the observation and background error covariances R and B.

We already estimate the observation bias correction coefficients (VarBC) and model error (via weak-constraint formulation) as part of the state estimation in 4D-Var. Could we include estimation of observation standard deviation into the state estimation problem? Similar to the bias correction, the observation errors could be parametrised based on some predictors. There could be a null space problem, with the assimilation having to choose between acting on or down-weighting innovations, but maybe more information would be needed to inform this decision. In this case the system could possibly be constrained using the metrological estimates.

4. Future Trends

4.1. New observations - small satellites, crowd sourced observations

For small satellites there are calibration challenges which may lead to errors with more complex structures, time evolution, and correlations in contrast to the current well calibrated, very stable platforms. They may only be available for short periods, so NWP centres will need to be quick in their data quality assessments. **A more automated system may be required for monitoring, quality control, bias correction, observation error specification for data from small satellites.** There would need to be a monitoring and alarm systems to alert to problems as part of any such automation. This seems to be an inevitable trend, so centres should start thinking about preparing for these small satellites now and this could lead to benefits in the assessment of all new satellites.

Observations of opportunity are often used to fill gaps that no existing observations can fill. They may not always be suitable for NWP e.g. river cameras for hydrology, car temperature gauges, which have complex observation operators and even more complex error structures.

4.2. Coupled assimilation

In weakly coupled systems the assimilation systems of the different components are separate. In more strongly coupled systems it is clear that the cross-correlations between components in the background errors are important. However, there may be implications for observation errors. The link is through representativeness error with observations in different components being sensitive to multiple components. For example, in a coupled ocean-atmosphere system there could be reduced representation errors due to better SST/emissivity as inputs to the atmospheric radiative transfer. Awareness of changes and new fields that could contribute to better radiative transfer and surface parameters is important. There was a suggestion that if coupling with an aerosol model, the outputs could be used in a situation dependent observation error model based on dust.

5. Priorities and final summary

The working group agreed on the following priorities for future work (all in bold in the main text of the report):

- Improved or alternative estimation techniques – More work on metrological estimates to attempt to reduce the amount of ad-hoc adjustments we need to make to estimates from the departure-based diagnostics.
- Understanding the influence of background and model error on diagnosed observation errors.
- More situation dependence and combining this with treatment of correlations. There is lots of potential for short-term improvements here, albeit with the caveat of increased sampling error when splitting the error estimates into different situations.
- Spatial error correlations
- More automated or online estimation of observation errors - required when dealing with lots of new satellite instruments simultaneously e.g. clusters of small satellites

Working group 4: Treatment of observation errors II

Chairs: Joanne Waller (Met Office), Stephen English (ECMWF)

Participants: Chris Burrows, Neill Bowler, Chawn Harlow, David Simonin, Mohamed Dahoui, Olaf Stiller, Anthony Weaver, Joe Nielsen, Sarah Dance, Niels Bormann, James Cotton, Leonhard Scheck, Oliver Guillet, Tijana Pfander, Yann Michel

1. Overview

This working group discussed the treatment of random observation errors in data assimilation, including both how uncertainties could be better estimated and accounted for in the assimilation schemes. The working group consisted of members from both operational and academic backgrounds, though all agreed that a better understanding of observation uncertainties is necessary to extract the maximum benefit of the observations in assimilation. The working group participants discussed a number of topics which are discussed in more detail below, but the three areas of research identified as priorities by the working group were:

1. Better understanding of the uncertainty estimation tools we have, and the estimates that they produce.
2. The increased usage of situation dependent observation uncertainties
3. Overcoming the technical challenges that currently limit the use of horizontal error correlations.

These priorities are given in no particular order as they are likely to be tackled by different research areas e.g. priority 3 is likely to be tackled first by those working in convective scale NWP, whereas priority 2 will be more important for global satellite assimilation.

An additional suggestion of this working group, though not a scientific research priority, was that increased collaboration and sharing of results may allow more rapid advancement in the treatment of observation errors. More regular meetings would allow centres to share their current progress with regards to the treatment of observation errors and would facilitate the exchange of new ideas and imminent research priorities. Therefore, it was suggested that it would be beneficial to hold more meetings such as this workshop.

2. Estimation tools and their accuracy

The group discussed a number of estimation tools that are currently available, though the majority of the discussion focused on the metrological approach, triple colocation methods and innovation based diagnostics. Despite the variety of tools available it was agreed that no one tool covered all our needs; the use of more than one method would be beneficial as comparison of results could increase confidence in our estimated observation uncertainty. For example, it was discussed that the most systematic methods, such as the metrological approach, are able to provide great insight in to the characteristics and sources of each uncertainty; however, this approach can be time consuming and may not be able to be updated rapidly or be applied to short lived instrumentation. Furthermore, the metrological approach gives rise to highly sophisticated error estimates that may have a high value, but to date have not often been used in an assimilation system, owing in part to their complexity. An extension of the metrological approach is to model the error from physical principles, using as far as possible traceable uncertainty in the modelling. This also requires physical modelling of the observation operator error. The triple colocation method was also discussed as a possible tool, though it was noted that this requires three independent sources of observation (or model) information, making it more applicable to some types of observations, e.g. surface observations like altimeter than others e.g. satellite radiances. Discussion regarding the Desroziers et al (2005) diagnostics, highlighted their ease of use but also the concern about the approximate nature of the resulting estimates. The use of such diagnostics is likely to continue in NWP, particularly for the routine monitoring of observation error statistics, and therefore it was decided that there should be more effort to understand the results of such diagnostics. This could be done using both theoretical investigations and by comparing observation error estimates obtained using different methodologies. Another statistical approach is the Hollingsworth-Lonnberg (1986) though it is used less widely than Desroziers. Where it is used it is generally used alongside Desroziers to increase confidence in the Desroziers estimate.

In summary the Desroziers approach is of great value due to ease of use and possibility of frequent update, including modelling observation error changes, but it needs to be used alongside other approaches that are more rigorous such as triple point, use of metrological data and physical modelling of observation error, the latter aiding understanding why statistically derived errors take the shape they do.

Recommendations

1. Develop the metrological approach for more observation types to provide insight in to error sources.
2. Routinely monitor observation uncertainties using innovation-based methods.
3. Provide confidence in routine estimates using sophisticated estimates.

3. Accounting for correlated observation errors: Current status, benefits, barriers and possibilities

Informed by members of the working group it was clear that the vast majority of operational centers are using inter-channel error correlations for hyperspectral and geostationary satellite observations (though there was less use for other observation types). It was clear that there is benefit to assimilating observation with these error correlations, though it was also noted that all-sky assimilation seems to work well with a diagonal R matrix; understanding why this is the case was deemed a priority.

Compared to inter-channel correlations, the use of spatial observation error correlations is much less prevalent; the only known use of horizontal error correlations was for Doppler radar winds at the Met Office. It was acknowledged that information is emerging regarding the benefit of accounting for horizontal error correlations, particularly in convection permitting NWP, though there are currently significant technical challenges that will need to be overcome if their use is to be more widely adopted. It was noted that, as well as the method developed at the Met Office for Doppler radar winds, there are also other methods that are being explored such as the potential use of diffusion operators and the assimilation of gradient observations. It was noted that each of these approaches would likely be applicable to different observation types. The challenge of simultaneously accounting for spatial and inter-channel correlations was also raised, and this should be considered alongside the development of the algorithms/infrastructure to account for horizontal error correlations in high resolution NWP.

Recommendations

1. Understand why a diagonal R matrix is so successful for all-sky assimilation.
2. Develop the algorithms/infrastructure to account for horizontal error correlations in high resolution NWP.

4. Situation-dependent observation uncertainties

All-sky systems already successfully use situation-dependent variances; the variance can be set either using cloud predictors or tuned using the outputs of a 1D Var pre-processing. The use of situation dependent inter-channel error correlations is an active area of research. However, the use of these correlations will also bring additional elements of complexity to the assimilation. Therefore, it will be important to understand the potential benefit of using these situation dependent correlations. It was noted that all-sky is not the only possibility for including situation dependence in the observation error statistics; though the 'situation' on which the errors may depend will be different for different observation types. It was agreed that the development of situation dependent observation error statistic for other observation types had potential.

Recommendations

1. Investigate the potential of situation-dependent inter-channel error correlations.
2. Explore the potential for other situation-dependent observation error statistics.

5. Additional Discussion

In addition to the three main priorities, the group also highlighted some other interesting, though lower priority, areas of research that had potential to advance the treatment of random observation errors in data assimilation.

The inclusion of correlated observation errors has been known to make the convergence of the assimilation slower, though it was noted that it is not always the case. If the convergence speed is impacted then it may be necessary to recondition the estimated observation error matrices by increasing the smallest eigenvalues of the R matrix. Operationally both the ridge regression and eigenvalue floor methods have been successfully used, but it was felt that there could be more research into the effects of these methods. In addition, there was also a desire to have a better mathematical and physical interpretation of small eigenvalues to ensure they are dealt with appropriately.

To date, temporal error correlations have received even less attention than spatial error correlations. There are a number of new instruments coming online that will produce observations with a high temporal resolution, and it is possible that for such observations the temporal correlations could become important. However, experience within the group suggested that increasing the observation frequency can be beneficial without accounting for the temporal correlations.

The fundamental assumption of the independence of background and observation errors was questioned, particularly for some types of derived observation, such as AMVs, but also for IR radiances for which cloud-screening relies on background departures. The group discussed if our current assimilation and error diagnostic tools would be appropriate if this assumption was relaxed. It was felt that there should be more research to understand this assumption, though this should be studied in simplified models in the first instance.

It was noted that in coupled earth system assimilation the current observation error statistics may no longer be a suitable representation as different model resolutions or new coupled boundary processes may alter the representation error associated with the observations. As we look to future assimilation systems it will be important to continue to refine our observation uncertainty estimates and to reassess the priority of those research topics discussed here.