Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

861

All-sky microwave radiances assimilated with an ensemble Kalman filter

Massimo Bonavita, Alan J. Geer, Mats Hamrud
(Research Department)

February 2020

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our website under:
http://www.ecmwf.int/en/publications

Contact: library@ecmwf.int

Abstract

Recent success in assimilating cloud and precipitation-affected satellite observations using the 'all-sky' approach is thought to have benefitted from variational data assimilation, particularly its ability to handle moderate non-linearity and non-Gaussianity, and to extract wind information through the generalised tracer effect. Ensemble assimilation relies on assumptions including linearity and Gaussianity that might cause difficulties when using all-sky observations. Here, all-sky assimilation is evaluated in a global ensemble Kalman filter (EnKF) system of near operational quality, derived from an operational four-dimensional variational (4D-Var) system. To get EnKF working successfully required a new all-sky observation error model (the most successful approach was to inflate error as a multiple of the ensemble spread) and adjustments to localisation. With these improvements, assimilation of 8 microwave humidity instruments gave 2% to 4% improvement in forecast scores whether using EnKF or 4D-Var. Correlations from the ensemble showed that all-sky observations generated sensitivity to wind, temperature and humidity. EnKF increments shared many similarities with those in 4D-Var. Hence both 4D-Var and ensemble data assimilation were able to make good use of all-sky observations, including the extraction of wind information. In absolute terms the EnKF forecast performance in the troposphere was still worse than the 4D-Var, though the gap could be reduced by going from 50 to 100 ensemble members. EnKF errors were larger in the stratosphere, where there are excessive gravity-wave increments that are not connected with all-sky assimilation.

1 Introduction

The assimilation of cloud and precipitation affected satellite radiances in the microwave part of the spectrum is one of the main success stories of global Numerical Weather Prediction (NWP) of the past 10 years (all-sky assimilation, Geer et al., 2018). Using observations affected by cloud and precipitation allows NWP centres to greatly increase the use of available satellite radiances and to consistently and significantly improve analysis accuracy and forecast skill (Geer et al., 2017). On the other hand, the assimilation of cloud and precipitation affected radiances challenges the assumptions made in standard data assimilation algorithms. First, the variables that we wish to analyse have physical bounds, and their errors depend on their proximity to these bounds (e.g. Posselt and Bishop, 2018; Bishop, 2019). Second, there typically is a highly non-linear relationship between the observed quantities (cloud and precipitation affected radiances) and the control vector variables (e.g. Errico et al., 2007; Bonavita et al., 2018) which also introduces significant deviations from Gaussianity in the prior and, to a lesser extent, in the posterior error distributions. Finally, and possibly most importantly, representation errors and model errors (when not separately accounted for) become the dominant error source for these observations (e.g. Geer and Bauer, 2011).

The most extensive use of all-sky radiances for operational forecasting is currently made at the European Centre for Medium-range Weather Forecasts (ECMWF), using the Integrated Forecasting System (IFS). This is based on hybrid four-dimensional variational data assimilation (4D-Var). 4D-Var uses the 'generalised tracer effect' to extract information that is not directly observed – for example it can extract wind information from observations sensitive to humidity and hydrometeors (Geer et al., 2017). Further, the variational approach enables solutions to many of the difficulties of using cloud and precipitation observations: the first problem (boundedness of the control variable and non-Gaussianity of its errors) can be tackled with a regularising transformation of the moist control variables that can include condensed water species (see Hólm et al., 2002; Geer et al., 2018), though extensions to a non-Gaussian variational framework have been put forward (e.g. Fletcher and Jones, 2014). The second problem, i.e. the non-linear relationship between observed and analysed quantities, can be treated by repeated re-linearisations of the variational minimisation problem around progressively more accurate first guess trajectory solutions (i.e. the 'inner-loops' in incremental 4D-Var, Courtier et al., 1994; Bauer et al., 2010; Bonavita

et al., 2018). The third problem (large errors of representation) is typically dealt with through the use of adaptive observation error models which take into account the degree of cloudiness present in the observations and the model (e.g. Geer and Bauer, 2011; Zhu et al., 2016; Migliorini and Candy, 2019).

Ensemble methods are a popular alternative to the variational approach: the Ensemble Kalman Filter (EnKF, e.g. Houtekamer and Zhang, 2016) is widely used in research; the ensemble-variational technique (EnVar) is used at some operational centres (e.g. Buehner et al., 2015; Kleist and Ide, 2015). The key difference in the EnKF and EnVar compared to 4D-Var is that background error covariances are propagated in time using ensemble correlations. In 4D-Var, this is done with tangent-linear (TL) and adjoint forecast models, which can be time-consuming to implement. Further, in currently used Krylov space minimisation algorithms (Gürol et al., 2014) it is necessary to run the TL and adjoint multiple times sequentially, which is less computationally scalable than running an ensemble of models in parallel. Currently, the best operational global forecasts come from hybrid assimilation systems, retaining the TL and adjoint approach alongside an ensemble component to provide situation-dependence in the background errors (Bonavita et al., 2016; Lorenc and Jardak, 2018). Nevertheless, ensemble approaches have been very popular for research on the assimilation of cloud and precipitation, for example in the explosion of work on all-sky infrared radiances in regional and local area models (Harnisch et al., 2016; Zhang et al., 2016; Minamide and Zhang, 2017; Honda et al., 2018; Zhang et al., 2018; Okamoto et al., 2019; Sawada et al., 2019).

From an ensemble perspective, the challenges of all-sky data assimilation are similar to those facing variational methods. However, there may be additional obstacles in assimilating all-sky radiances: vertical covariance localisation may be inappropriate for an observation that is sensitive to cloud at any level in the troposphere (e.g. Houtekamer and Zhang, 2016; Geer et al., 2018; Okamoto et al., 2019); the required covariance inflation could depend on the presence of cloud and precipitation (e.g. Minamide and Zhang, 2019). More fundamentally, an ensemble method like the EnKF relies on a local solver, a Gaussian statistical framework and a linear analysis update. Further, although ensemble assimilation has a clear theoretical ability to extract winds from humidity and even from hydrometeors (e.g. Allen et al., 2015; Lien et al., 2016) this has not been demonstrated practically in an operational-quality system with all-sky radiance observations.

Here, we compare the impact of all-sky microwave radiances on analysis and forecast skill in EnKF and 4D-Var (non-hybrid) versions of the IFS. Hence this study extends the results of Hamrud et al. (2015) and Bonavita et al. (2015), but with the addition of all-sky radiance assimilation, which was missing from the EnKF they tested, but not from the 4D-Var. We will investigate whether all-sky assimilation can be successful using an ensemble method, despite the stricter need for linearity and Gaussianity. It is also of interest whether the addition of all-sky observations would have made the EnKF version of the IFS more competitive with the 4D-Var version.

In this paper, Section 2 briefly introduces the EnKF and 4D-Var versions of the IFS, the all-sky microwave observations coming from 8 imaging and humidity-sounding sounding instruments, and the set of experiments that we have used. Observation error modelling is key to all-sky data assimilation, but it is not obvious how best to transfer the existing Geer and Bauer (2011) approach into an ensemble framework. Hence section 3 covers this aspect in detail (but can be skipped by those with less interest). This section introduces two new approaches to observation error modelling that may be more appropriate for ensemble data assimilation, and explores how to compute data assimilation error budgets in an ensemble context. Section 4 gives the results: it first examines the ensemble sensitivities to the all-sky observations and explores the best EnKF configuration, considering different possibilities for localisation and observation error modelling. Then the impact of all-sky assimilation is assessed in the EnKF and 4D-Var contexts, looking at both increments and forecast scores. A final subsection examines the

absolute difference between EnKF and 4D-Var, now that we have consistent observation usage between the two.

2 Method

2.1 EnKF version of the IFS

The Ensemble Kalman Filter used in this paper is a local ensemble transform Kalman filter (LETKF, Hunt et al., 2007) and its ECMWF implementation has been described in Hamrud et al. (2015). The experiments have been run at TCo319 resolution (triangular truncation at 319 spectral components over a cubic octahedral grid, corresponding to approx. 39km grid spacing), 137 vertical levels ranging from near surface to 1 Pa, with a 6 hour assimilation window and 50 ensemble members. The IFS model version used in all experiments is cycle 42r1, which incorporates only minor technical changes with respect to the IFS cycle used in the experiments described in Hamrud et al. (2015). As described in that paper, covariance inflation and covariance localisation are heuristic but necessary measures to improve the performance of EnKF-based assimilation. In our experiments we have used both multiplicative covariance inflation (relaxation to prior variance, after Whitaker and Hamill (2012), with a 0.9 relaxation factor) and additive covariance inflation (from a climatology of rescaled 24 h forecast differences).

Covariance localisation is commonly used in the EnKF to combat the effects of the rank deficiency of the sampled background error covariances (e.g. Houtekamer and Zhang, 2016). In the LETKF used here, calculations are done in ensemble space. Hence we use a combination of domain localisation, where a fixed number of observations are selected in a local domain around the gridpoint being analysed (see Hamrud et al., 2015) and **R** localization (Hunt et al., 2007; Greybush et al., 2011) where the observation errors of the selected observations are inflated according to the distance from the analysed grid point. While conceptually different from the standard covariance localization which is applied to the **B** matrix or its projection in observation space ($\mathbf{HBH}^T$), results are found to be comparable (Houtekamer and Zhang, 2016, and references therein). A benefit of **R** localization, is that different localization parameters can be used, appropriate to the horizontal and/or vertical support of different types of observations. 2000 km is the horizontal distance of zero impact of an observation on an analysis grid point, while vertical localization is over 2 scale heights (zero impact vertical distance). Changes to these baseline values were also tested as part of the experimentation and will be described later.

2.2 4D-Var version of the IFS

An adapted version of the 4D-Var IFS is used as a reference with which to evaluate the performance of the EnKF. The previous experiments of Bonavita et al. (2015) had limitations: first, that the EnKF experiments did not use all-sky radiances, while the reference 4D-Var did; second, the 4D-Var configuration was that of the operational Ensemble Data Assimilation (EDA) of the time, which employed only two outer loop relinearisations. As shown more recently, when using all-sky observations in the ECMWF 4D-Var, more relinearisations lead to improved performance (Bonavita et al., 2018).

The new incremental 4D-Var reference experiment shares the same outer loop and forecast model resolution as the EnKF experiments described in this paper (TCo319, 137 vertical levels) but it runs three minimisations at TL159/ TL191/ TL255 resolution (triangular truncation at 159/191/255 spectral components over a linear reduced Gaussian grid, corresponding to approx. 120/100/80 km grid spacing). While these resolutions are smaller in an absolute sense than those used in the operational 4D-Var

Satellite	Sensor	Channels [GHz, polarisation]	Utilisation
Metop-A	MHS	183±1, 183±3, 190.3	Ocean, snow-free land, snow-covered land, sea-ice
Metop-B	MHS	183±1, 183±3, 190.3	”
NOAA-18	MHS	183±1, 183±3, 190.3	”
NOAA-19	MHS	183±1, 183±3, 190.3	”
GCOM-W	AMSR2	19v, 19h, 24v, 24h, 36v, 89v	Ice-free ocean between 60°N and 60°S
GPM	GMI	19v, 19h, 24v, 24h, 36v, 89v	”
DMSP F-17	SSMIS	19v, 19h, 22v, 37v, 92v, 183±1, 183±3, 183±7	Low frequency channels as AMSR2; high frequency channels as MHS
DMSP F-18	SSMIS	183±1, 183±3, 183±7	Ice-free ocean

Table 1: All-sky observations added in the EnKF and 4D-Var experiments in this work. The lowest peaking WV sounding channels of MHS, 190.3 GHz and 183±7 GHz, are used like the AMSR2 imager channels, i.e. not over land or sea-ice. Polarisations (v=vertical, h=horizontal) are conventionally specified only for imaging channels.

(TL255/TL319/TL399), in relative terms this is a skillful incremental 4D-Var setup for this model resolution as the ratio of the outer to inner loop resolution is only about 2 (compared with a ratio of approx. 5 for the operational 4D-Var). Apart from the resolution there are three other significant differences with respect to the operational 4D-Var. One is the assimilation window length, which is set to 6 hour instead of the 12 hour of the standard ECMWF 4D-Var. This choice was motivated by the desire to have a set-up of the data assimilation cycle as close as possible to the one used for the EnKF in order to make comparison of the algorithms easier. In terms of analysis and forecast skill, the 6 h cycling results in only minor performance differences. Another consequence of the 6-hour cycling choice and the unavailability of a corresponding 6-hour cycled EDA (Isaksen et al., 2010) is that 4D-Var was run using climatological background error covariance estimates. Based on recent tests, the use of a static B is expected to degrade 4D-Var performance by approx. 2% in standard tropospheric performance measures. The final difference consists in the fact that current operational 4D-Var runs with four outer loop re-linearisations instead of the three used in the experiments described in this work. Based on the results of Bonavita et al. (2018), this is expected to degrade 4D-Var performance by approx. 1%.

2.3 All-sky microwave observations

The IFS assimilates a suite of all-sky microwave radiance observations that is now large enough to provide around 20% of all forecast impact at 24 h as measured using an adjoint-based sensitivity diagnostic (Forecast Sensitivity to Observation Impact, FSOI, Langland and Baker, 2004; Cardinali, 2009). The all-sky observing system used in the IFS is described in detail by Geer et al. (2017), along with the aforementioned FSOI results. Compared to the cycle 41r1 configuration described in that work, the cycle 42r1 configuration additionally has all-sky water vapour sounding (WV, 183 GHz) channels actively assimilated over snow-covered land surfaces, along with the addition of the Special Sensor Microwave Imager Sounder (SSMIS) F-18 WV sounding channels over ocean. Table 1 summarises the set of all-sky observations used in the current work. See the OSCAR website (`https://www.wmo-sat.info/oscar/`) for full descriptions of the satellites and their acronyms.

Further methodological details of the all-sky microwave assimilation are given by Bauer et al. (2010); Geer et al. (2010); Geer and Bauer (2011). The implementations of WV sounding channels of SSMIS and the Microwave Humidity Sounder (MHS), Advanced Microwave Scanning Radiometer 2 (AMSR2)

Name	Expt. ID	All-sky	Localisation	Ensemble members	Obs. error
EnKF baseline	ghkr	Off	1	50	N/A
EnKF all-sky localisation 1	ghkm	On	1	50	Symmetric
EnKF all-sky localisation 2 / symmetric	gjaq	On	2	50	Symmetric
EnKF all-sky nonlinearity	gi67	On	2	50	Nonlinearity
EnKF all-sky spread-multiple (50)	gj6h	On	2	50	Spread multiple
EnKF all-sky spread-multiple 100	h5hf	On	2	100	Spread multiple
4D-Var baseline	h58p	Off	N/A	N/A	N/A
4D-Var all-sky	h542	On	N/A	N/A	Symmetric

Table 2: Experiment summary

and GPM microwave imager (GMI) are described by Geer et al. (2014), Kazumori et al. (2016) and Lean et al. (2017) respectively. In addition to the all-sky microwave data, the observations assimilated in the IFS include: microwave sounders not yet converted to all-sky assimilation, notably the microwave temperature sounders; infrared sounders on polar orbiting and geostationary satellite platforms; scatterometer surface winds; global navigation satellite system (GNSS) bending angles and in-situ and/or ground-based measurements from aircraft, radiosondes, ships, buoys and surface stations. This observing system is comprehensive enough that even when the all-sky microwave observations are removed, the quality of medium-range forecasts deteriorates by only around 3 – 4% (Geer et al., 2017). However, this is still about as large an impact as can be obtained by removing one major component of the global observing system (compare e.g. Bormann et al., 2019) and the extensive use of all-sky microwave observations is a key advantage in the overall quality of ECMWF forecasts.

2.4 Experiment summary

Experiments were run in the EnKF and 4D-Var configurations as summarised in Tab. 2. In each case, a baseline experiment was created by assimilating the full observing system minus the all-sky microwave sensors listed in Tab. 1. These observations were then added back in to create a number of all-sky experiments. The 4D-Var all-sky configuration is that of the operational IFS at cycle 42r1, apart from the changes in resolution and length of the assimilation window described earlier. Various configurations of localisation, observation error model, and ensemble size have been tested in the EnKF all-sky experiments.

The EnKF experiments ran from 31 July to 31 October 2015, giving a maximum of 93 long forecasts for verification purposes. The exception was the experiment with the original covariance localisation, which only ran until 1 October 2015, after which it was cancelled due to poor results and to save resources. The 4D-Var experiments were run for one day less, from 1 August to 31 October.

3 All-sky observation errors for ensemble assimilation

One reason that all-sky assimilation was not available in the earlier EnKF experiments (Hamrud et al., 2015; Bonavita et al., 2015) was the difficulty of implementing the observation errors. Errors become larger in the presence of cloud and precipitation due to inaccuracies in fast modelling of scattering radiative transfer, errors of representation, and finally model error (both the lack of predictability, and the presence of state-dependent systematic model errors particularly in cloud and precipitation). Model error is not explicitly represented in the troposphere in the IFS 4D-Var or in our implementation of the EnKF, so it must be included as part of the observation error.

To model the effect of cloud on observation error, Geer and Bauer (2011) inflated observation errors as a function of a cloud predictor. A cloud proxy variable c is inferred using a simple retrieval from observation space (any y) to be represented here as $c = g(y)$. The original cloud retrieval, used for the microwave imagers SSMIS, AMSR2 and GMI over oceans, uses the normalised polarisation difference at 37 GHz, which is sensitive to water cloud and rain absorption. Over land for SSMIS and for the microwave humidity sounders (MHS) over all surfaces, where the sensitivity is to scattering from frozen particles, one of two scattering indices is used (one for ocean, one for land, Baordo and Geer, 2016; Geer et al., 2014). The cloud amount is made 'symmetric' by taking the average of the cloud amount estimated from the observations y^o and simulated from the model background $H(x^b)$ using the nonlinear observation operator $H()$:

$$\bar{c} = \frac{1}{2}(c^o + c^b) \equiv \frac{1}{2}\left(g(y^o) + g(H(x^b))\right). \tag{1}$$

This is a good predictor for the standard deviation of background departures. The background departure standard deviations are binned as a function of symmetric cloud amount, and these values can either be put in a lookup table (e.g. Okamoto et al., 2014; Harnisch et al., 2016) or fitted using a piecewise linear or quadratic fit (e.g. Geer and Bauer, 2011; Geer et al., 2014). This fit is then used as a model to predict the observation errors, possibly after subtracting an insignificant amount of variance to represent clear-sky background errors. Hence this approach assumes that observation errors dominate the total error.

As well as inflating the error variances, the presence of cloud increases inter-channel observation error correlations (Bormann et al., 2011; Okamoto et al., 2019). However the combination of inter-channel observation error correlations with an all-sky error inflation technique has only just started development (Geer, 2019). Hence in this work, observation error correlations are ignored. We simplify the following equations using scalar notation, but the full matrix-vector versions could easily be substituted (see e.g. Dee, 1995; Hunt et al., 2007).

For an ensemble data assimilation system implementing the framework of Geer and Bauer (2011), it is not immediately clear how to represent the increase in observation error coming from the presence of cloud in the background ensemble. With i the index of an ensemble member and the overbar representing an ensemble average, candidates for the model cloud proxy include:

1. the ensemble mean in observation space $g\left(\overline{H(x_i^b)}\right)$ as in Harnisch et al. (2016);
2. the ensemble mean in model space $g\left(H(\overline{x_i^b})\right)$ as in Minamide and Zhang (2017);
3. the mean of cloud retrieved from different members $\overline{g\left(H(x_i^b)\right)}$.
4. the control $g(H(x_{ctrl}^b))$.

We used the control member (option 4) in our testing because it was the most practical. However the presence or lack of cloud in the control member is not a perfect predictor of the presence of cloud in any of the other ensemble members. The other options highlight that introducing nonlinearity into the EnKF framework relies on careful linear approximation (Hunt et al., 2007). In this work the cloud retrievals $g()$ are linear functions of the observations, so option 3 is identical to option 1, but nonlinear cloud retrievals can be encountered (e.g. Zhu et al., 2016). To summarise, when we tested the symmetric error approach in the EnKF we used all the same settings as for 4D-Var, except the choice of the control member to represent the model cloud amount.

The symmetric error approach has been successful in both 4D-Var (Geer et al., 2017) and EnKF (e.g. Okamoto et al., 2019) but a disadvantage is that it may violate Bayes theorem by using knowledge from the observation in order to estimate the likelihood. With limited time for this project, we explored two new candidate all-sky observation error models that fit naturally into the EnKF system, rather than trying to further develop the symmetric error approach, or to test other possible candidates (e.g. Minamide and Zhang, 2017). The two new candidates use information from the ensemble, not the observations, which means Bayes is better respected.

The first new candidate is the 'nonlinearity' error model. It inflates the observation error above a clear-sky minimum error $\sigma^o_{clear-sky}$ as a function of the difference in brightness temperature between the control and the ensemble mean:

$$\sigma^o = \sigma^o_{clear-sky}\left(1+\alpha\frac{\left|H(x^b_{ctrl})-\overline{H(x^b_i).}\right|}{\sigma_o^{clear-sky}}\right) \tag{2}$$

The divergence between the control (initialised from the ensemble mean analysis) and the mean of the ensemble members (initialised from the previous analysis plus a perturbation) is a measure of the combined nonlinearity of the 6 h forecast and the observation operator. Most of the nonlinearity diagnosed by Eq. 2 likely comes from cloud and precipitation (Bonavita et al., 2018). The parameter α is a tuning factor. With $\alpha = 1$ the model assumes that nonlinearity error is the only source of additional observation error in cloudy conditions. If $\alpha > 1$ it can allow additional error sources (such as representation error) as long as they also depend on the size of the nonlinearity. We started with $\alpha = 1$ but this gives relatively small observation errors in cloudy situations compared to the other candidates (see later).

The second candidate error model, the 'spread multiple', uses the ensemble spread to inform the observation error. Satterfield et al. (2017) have shown that ensemble variance can be a good predictor of the error of representation. In reverse, Minamide and Zhang (2019) have shown how all-sky departures can drive an adaptive background spread inflation in an EnKF. Harnisch et al. (2016) compared the spread of an ensemble Kalman filter with all-sky infrared background departures, binned as a function of the symmetric cloud predictor. The shape of the two curves was similar but the background spread was around 3 times smaller in variance terms. If observation error can be represented as a multiple of the ensemble spread variance, and making sure to keep the clear-sky observation error as a minimum floor, it would be described as follows:

$$\sigma^o = max\left(\sigma^o_{clear-sky}, \left(\beta Var(H(x^b_i))\right)^{\frac{1}{2}}\right). \tag{3}$$

Here $Var()$ is an operator for computing variance over the ensemble (over all i). β is the multiplication factor, with $\beta = 3$ the initially chosen value, in a cautious approach giving observation errors that are large relative to the background departure standard deviations (see later).

The assumed errors can be tested in the ensemble framework by comparing the sum of background spread and observation error variances to the background departure variance. In deterministic data assimilation the standard error analysis (Dee, 1995; Desroziers et al., 2005) has no ambiguity over how to compute

the background departures. But with an ensemble it is not obvious whether to use departures from the ensemble mean, members or control to compute the error budget. Some ensemble studies have not recorded which version they used and others used the ensemble mean but without detailed justification (e.g. Houtekamer et al., 2005; Harnisch et al., 2016). Hence we devote some space to explaining how the error budget can be correctly computed in an ensemble context.

In the LETKF (Hunt et al., 2007, Sec. 2.2.1) the ensemble mean represents the background state, which would have an error ε^b from the true state of the atmosphere x^t:

$$\overline{x^b} = x^t + \varepsilon^b \tag{4}$$

The unknown statistics of ε^b give the true PDF and covariances of background errors. To estimate of the background error covariance, the LETKF uses the background ensemble spread $Var(x_i^b)$. Each member i of the background ensemble can be seen as representing the ensemble mean plus an ensemble perturbation η_i^b:

$$x_i^b = \overline{x^b} + \eta_i^b \tag{5}$$

The ensemble perturbation η_i^b is intended to be another draw from the PDF of true background error but this depends on the quality of the LETKF implementation. To complete our survey of errors, an observation has a true observation error ε^o, here for simplicity assuming the observation operator $H()$ is perfect:

$$y^o = H(x^t) + \varepsilon^o. \tag{6}$$

The departure from each ensemble member can be defined and expanded as follows, assuming local linearity and introducing $\mathbf{H}$ as a linearised version of the observation operator:

$$y^o - H(x_i^b) \simeq \varepsilon^o - \mathbf{H}\varepsilon^b - \mathbf{H}\eta_i^b. \tag{7}$$

Using $E()$ as the expectation operator, the expected value of the ensemble member departure variances can be simplified as follows, assuming there are no correlations between background, observation error, and ensemble perturbations:

$$E((y^o - H(x_i^b))^2) \simeq E((\varepsilon^o)^2) + E((\mathbf{H}\varepsilon^b)^2) + E((\mathbf{H}\eta_i^b)^2) \tag{8}$$

Hence the ensemble member departure variances should be the sum of the true observation error variance, the background errors and ensemble spread in observation space. Derived similarly, the expectation of the ensemble mean departures is the sum of the true error variances of observation and background:

$$E((y^o - \overline{H(x_i^b)})^2) \simeq E((\varepsilon^o)^2) + E((\mathbf{H}\varepsilon^b)^2)) \tag{9}$$

Hence the validity of the assumed errors can be tested by comparing the variance of the ensemble mean background departures (an estimate for the LHS of Eq. 9) to the sum of the assumed observation error variance and the background spread in observation space. Over a sufficient sample these should be equal if the errors are correctly specified. A scalar version of the χ^2 test can also be applied by computing the ratio of the ensemble mean departure to its predicted ('total') error standard deviation:

$$\frac{y^o - \overline{H(x_i^b)}}{((\sigma^o)^2 + Var(H(x_i^b)))^{\frac{1}{2}}} \tag{10}$$

Further assuming Gaussianity, this 'normalised departure' should, if the errors are well-specified, be distributed as a Gaussian with a variance of 1.

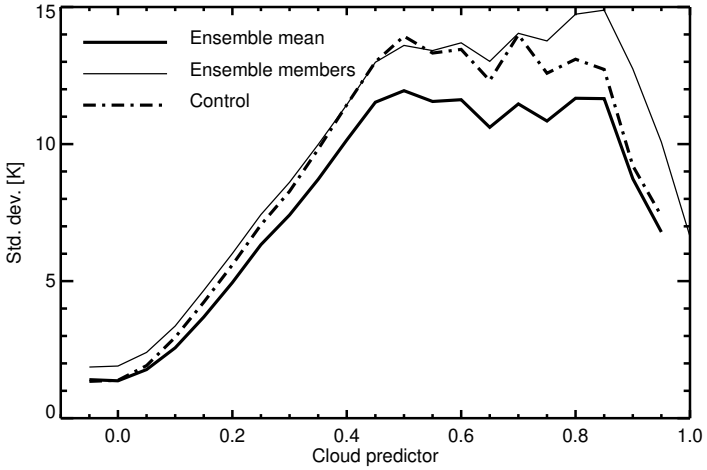

Figure 1: Standard deviation of the background departures computed against the ensemble mean, control and individual members, binned as a function of cloud proxy variable computed from the control member, using a sample of assimilated SSMIS F-17 channel 19v brightness temperatures from 1 - 10 September 2015 from the Symmetric error experiment

Based on a 10-day sample of assimilated observations, Fig. 1 bins the standard deviation of background departures as a function of the symmetric cloud amount (eq. 1, using $g(H(x_b^{ctrl}))$). As also seen by Harnisch et al. (2016), and consistent with Eqs. 8 and 9, the ensemble mean departures have smaller error standard deviations than those of the ensemble members. The control member departures $y^o - H(x^b_{ctrl})$ are also shown. For a linear model and observation operator, the control departure should be identical to the ensemble mean departure, as seen in clear-sky situations (identified by cloud predictor amounts around zero). But for cloud amounts between 0.4 and 0.8, the control member departures are closer to those of the individual members, illustrating the effect of nonlinearity. Beyond cloud amounts of 0.8, there are relatively higher values of the ensemble member departures for cloud amounts. This is likely due to the use of the control member cloud amount in the binning (since the presence of cloud in the control is not a perfect predictor of cloud in any ensemble member). We use the ensemble mean departure in the following assessment of the error models, since it is more theoretically correct. However, Eq. 9 assumes linearity and Gaussianity, so especially for all-sky assimilation, i.e. in nonlinear situations, we should be open to using larger errors.

Figure 2 shows PDFs of normalised ensemble mean departures (Eq. 10) using the different error models for channel 19v of SSMIS over the 10-day sample of assimilated observations. The PDFs are more peaked than the expected Gaussian, showing that modelled observation errors are large in comparison to those expected in a perfect linear and Gaussian ensemble framework. There are 'warm tails' to two of the PDFs. Channel 19v is sensitive to heavy rain with additional sensitivity to column water vapour and cloud. Adding rain makes the brightness temperature warmer, so the warm tails contain situations where the observation is likely cloudy or rainy but the ensemble is clear. To generate large positive normalised departures, likely the observation errors are too small because the ensemble does not contain sufficient precipitation to generate either nonlinearity (Eq. 2) or a large background spread (Eq. 3). Here, the symmetric error model has an advantage, since it takes into account the presence of cloud in the observation in order to boost the error (Geer and Bauer, 2011). The PDF of this error model is more symmetric.

Figure 3 shows the standard deviation of the ensemble mean departures ($y^o - \overline{H(x_i^b)}$), estimated back-

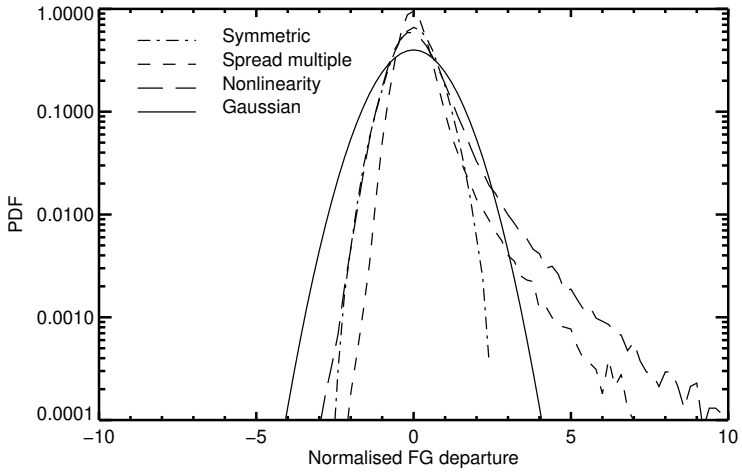

Figure 2: PDFs of ensemble mean departures normalised by estimated total error for the three different observation error models, using a sample of assimilated SSMIS F-17 channel 19v brightness temperatures from 1 - 10 September 2015 from the Symmetric error experiment. With an accurate combination of background error and observation error modelling, the normalised departures would have a Gaussian distribution (solid line)

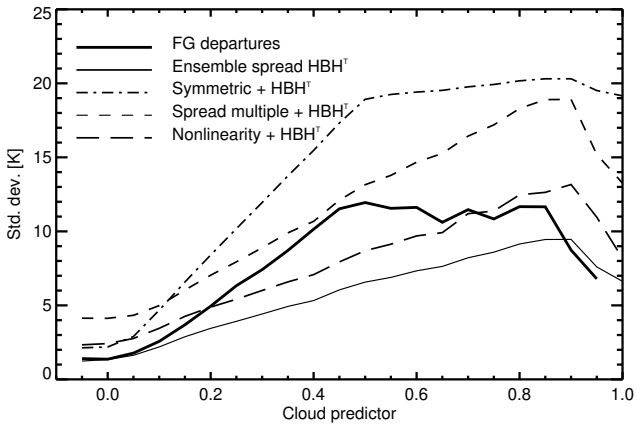

Figure 3: Standard deviation of the ensemble departures, background spread, and total error models, binned as a function of cloud proxy variable computed from the control member (sample as in Fig. 2): SSMIS F-17 channels (a) 19v and (b) 183±1 GHz

ground error from the ensemble spread in observation space, and the predicted total error (denominator of Eq. 10), as a function of the symmetric cloud amount. As expected the ensemble spread and the total error increase with the symmetric cloud amount, but dip again towards the maximum symmetric cloud amount, which requires cloud to be present in both background and the observation (Geer and Bauer, 2011; Harnisch et al., 2016). The symmetric error model gives a high estimate of observation error in more cloudy situations because it is derived from the total error assuming background errors are negligible (Geer and Bauer, 2011; Geer, 2019). Under this philosophy, the background error is included twice in the total error.

For channel 19v (panel a), the nonlinearity error model appears to underestimate the total error for mid-range cloud amounts (this likely corresponds to the observations forming the warm tail in Fig. 2). However for 183±1 GHz the nonlinearity model is relatively consistent with the background departures for cloudy scenes. The spread multiple approach produces a similar distribution of total error as a function of cloud, but with β=3 it gives relatively high total error. The results in most other channels follow the pattern of channel 19v. Channel 183±1 GHz is an outlier because it is sensitive to the upper-troposphere and its errors are as much driven by errors in humidity as in cloud and precipitation. In this initial testing, we have chosen fixed values of α and β for all channels; in practice these should vary as a function of channel like the parameters of the symmetric error model.

With the initial settings, and judged against the ensemble mean departures, the symmetric error model gives reasonable weight to clear-sky observations and too little weight to cloudy observations. The nonlinearity error model appears to give reasonable weights to both clear-sky and cloudy scenes. Finally the spread multiple gives apparently too little weight to clear-sky and cloudy observations. Judged against the ensemble member or control departures (such as illustrated in Fig. 1) the error models do not look so inflated (not shown): the nonlinearity error model would produce an underestimate of total error in cloudy scenes, and the other two models look more appropriate. Further, theoretical estimates of observation error typically need ad-hoc inflation to get good results in real assimilation systems, which may come from the many other sub-optimalities and assumptions used in data assimilation (Bormann et al., 2016). Further, we are relying on the background spread being a good estimate of the true background error. Therefore it might not be surprising if the apparently 'too large' errors work best in practice.

4 Results

4.1 Ensemble correlations and localisation

Figure 4 shows the correlations between model variables and satellite radiance observations across the EnKF 50-member ensemble, binned as a function of distance. Strictly this shows the RMS of the correlation, since the sign of the correlations is not always consistent; note also statistics were gathered in Fisher Z-transform space. The lower limit is 0.142 which is a floor set by random noise. The top row shows the correlations for AMSU-A channel 5, one of the most influential microwave channels assimilated in the clear-sky framework. The correlations span the entire troposphere and are even deeper than the channel's weighting function, which gives the direct temperature sensitivity of the instrument (see e.g. Lawrence et al., 2018). The additional correlations are thus geophysical (for example there is no direct sensitivity to wind, but this arises through geostrophic balance, among other causes). The middle row shows SSMIS channel 11, at 183±1 GHz, which is directly sensitive to upper-tropospheric water vapour and frozen precipitation. Interestingly this channel has stronger wind correlations than does AMSU-A, but they are more localised. The correlations with all variables span only the upper half of the tropo-

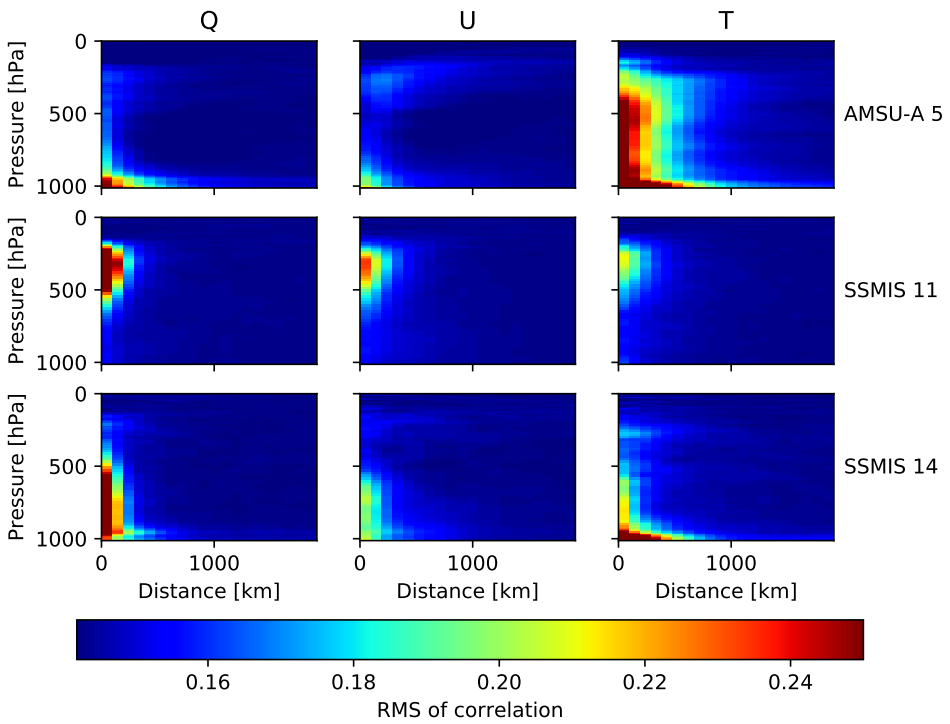

Figure 4: RMS of correlations of the background forecast variables (columns: Q= specific humidity; U=zonal wind component; T=temperature) with observations from AMSUA channel 5, SSMIS channels 11 and 14 (rows). Correlations computed from the 00 UTC cycle on 15th August 2015 in the 50 member spread-multiple configuration, and represent the northern hemisphere (North of 20°). The shading colours have been allowed to go off-scale to concentrate on the more distant correlations, but peak temperature correlations of AMSU-A are around 0.27 in the mid-troposphere, and peak SSMIS humidity correlations are around 0.4.

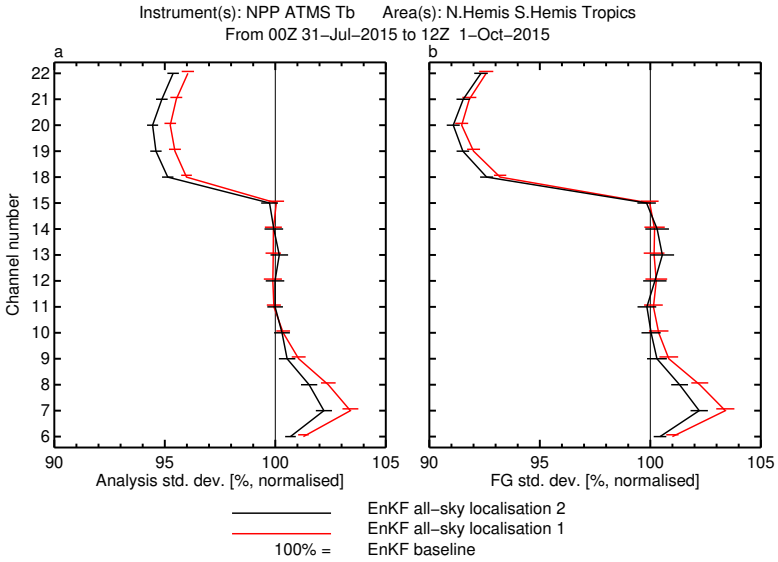

Figure 5: Normalised std. dev. of analysis and background departures from assimilated ATMS observations, showing, impact of all-sky assimilation in LETKF using two different versions of the vertical covariance localisation.

sphere, and extend only around 200 km from the observation, in contrast to around 500 km for AMSU-A channel 5. The lowest row shows SSMIS channel 14, at 22 GHz, which is directly sensitive to the surface, lower-tropospheric humidity, water cloud and rain. Similarly it has restricted spatial correlations of around 200 km horizontally, and up to around 500 hPa in the vertical. However there is some correlation higher up in the troposphere, likely associated with deep precipitating structures. There are more distant correlations in temperature near the surface, which likely comes from the sensitivity of the radiances to the skin temperature, which is then geophysically correlated with boundary layer temperatures. At the broadest level, the ensemble correlations reveal that all-sky observations made in primarily water-vapour channels bring a different kind of information compared to satellite temperature sounding: the information is more correlated across all variables including wind, but it is more restricted in the vertical and the horizontal.

In terms of localisation, these results suggest it is beneficial to have smaller localisation values for the all-sky radiances. Supporting this, Okamoto et al. (2019) have found better forecast quality in a mesoscale data assimilation system after decreasing the horizontal localisation radius for all-sky infrared radiances (to 100 km from 200 km). In our global system we tested the application of new settings to the all-sky observations: halving of the horizontal localisation compared to the default value (1000 km instead of 2000 km) and a reduction of 20% of the vertical localisation (1.6 scale height instead of 2).

Figure 5 shows the impact of all-sky assimilation, compared to the no-allsky baseline, using the two different localisations (1=original; 2=reduced scales for all-sky). Impact on forecasts is measured by the background fits to clear-sky ATMS observations. Note that in all the plots of this kind shown, for the ensemble experiments they are based on the departure from the control member. Our first all-sky experiment used the same localisation as other observations and produced disappointing results, particularly the degradation in fit to ATMS channels 6 – 10, which have temperature sensitivities spanning the troposphere and lower stratosphere (see e.g. Lawrence et al., 2018). Going to the 2nd localisation approach improved fits to ATMS by around 1% and was significantly beneficial, both for the tropospheric temperature quality but also the mid- and upper-tropospheric humidity, as measured by ATMS channels 18 – 22. ATMS channels 11 – 15 are unaffected by all-sky assimilation, which is to be expected as these

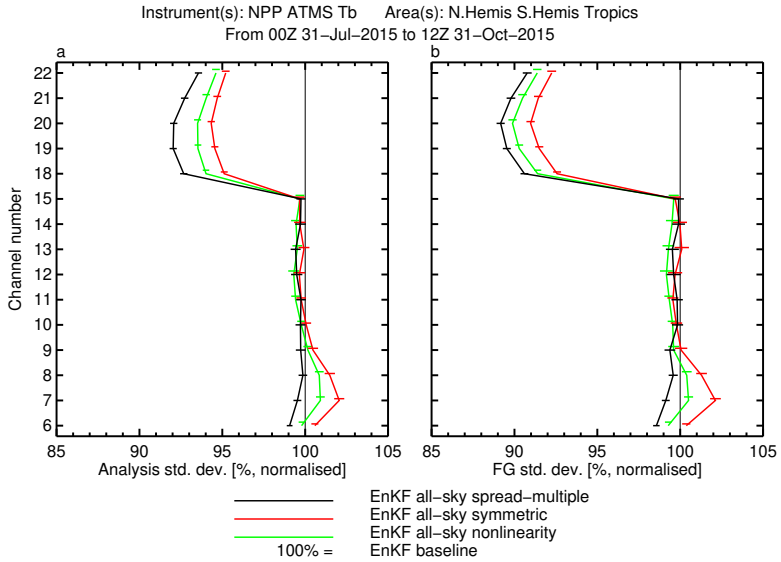

Figure 6: As Fig. 5 but showing three different observation error models

sound temperature in increasingly higher levels through the stratosphere. Other observational fits confirmed the beneficial impact of tighter horizontal and vertical localisation, but the experiments were not run long enough to establish statistical significance in the longer-range forecast verification. Even with the improved localisation the addition of all-sky assimilation still caused degradations in the background fits to tropospheric temperature observations. Hence we looked for another source of improvement by trying different types of all-sky observation error model.

4.2 Observation error model

The candidate observation error models were tested with the benefit of the improved localisation approach (i.e. reduced localisation lengths both horizontally and vertically). As measured by the fit to ATMS radiances (Fig. 6) both the 'nonlinearity' and the 'spread multiple' gave significantly better results than the symmetric error model. In particular the spread multiple allows all-sky assimilation to improve fits to the ATMS tropospheric temperature channels 6 – 9, whereas the other candidates allow these to degrade, particularly so in the case of the symmetric error model. The spread multiple error model is the only model to consistently improve, rather than degrade, the analysis fit. This cannot be interpreted as a measure of the overall quality of the analysis but if the fit is worsened, it could be due to additional perturbations being added into the analysis, perhaps if some observations are overfit because their errors are imperfectly specified. However, the background fits (panel b) are mostly improved with all three error models. Consistent results are seen with other observation types as the reference. As discussed earlier, none of the three error models has seen any tuning to get the best results in the EnKF, due to limited time to spend on this project and the computational cost of running the EnKF. Judged by a χ^2 statistic (not shown) the 'nonlinearity' model could have achieved similar weights as the spread multiple model, as a function of cloud amount, with its α parameter around 1.5 rather than 1.0.

Figure 7 shows verification of longer-range forecasts, using the experiments' own analyses as the reference. Confidence intervals are specified at 95% following the approach described in Geer (2016) with additional Šidák inflation for multiple testing. There is no statistically significant difference between the three experiments at day 2 or beyond, but the symmetric error model still generally gives the worst scores

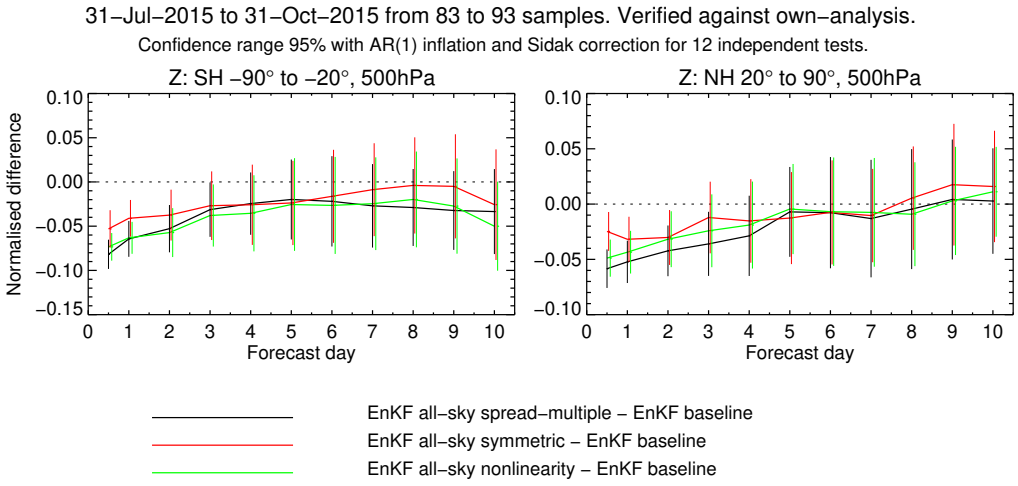

Figure 7: Normalised change in RMS error in 500 hPa geopotential height compared to a control without all-sky assimilation, using one of three proposed error models.

of the three, consistent with the observation fits. All three all-sky experiments improve the forecast out to at least day 3 and by up to 5%. Based on the short-range fits to observations, the spread-multiple error model was chosen for further testing.

4.3 Increments in LETKF and 4D-Var

To see how the EnKF and 4D-Var make use of all-sky observations, we examined the increments for a single analysis, 00 UTC on 15th August 2015. Three additional single-cycle experiments were run in which only the all-sky observations from Tab. 1 were assimilated; all other observations were excluded. Geer et al. (2014) used experiments like this to reveal strong similarities between wind increments made with a set of all-sky observations and with a full observing system. This demonstrated the ability of 4D-Var to extract wind information from all-sky microwave observations sensitive to water vapour, cloud and precipitation. Here, two single-cycle experiments used the 4D-Var control configuration, and one the EnKF configuration with 50 members (as used in the equivalent full-observing-system experiments listed in Tab. 2). The two single-cycle 4D-var experiments differed by their initial conditions, which in one were the background forecasts from the corresponding full-system 4D-Var, and in the other from the full-system EnKF. For the single-cycle EnKF experiment, the initial conditions came from the full-system EnKF. As will be seen the choice of initial conditions has a strong influence on the similarity of the increments between different experiments.

Figure 8 shows the increments in wind divergence at 200hPa from the two all-sky only experiments (both using the EnKF-derived initial conditions) along with the equivalent full observing system experiments at that time (see Tab. 2). For the 4D-Var experiments, the 'true' increments are made in the control vector at 21 UTC at the beginning of the window, but what is plotted is the difference between the updated (3h) and background (9h) forecast valid at 00 UTC. For EnKF, the plotted increments are the difference between the ensemble mean analysis and the control (6h) forecast, again valid at 00 UTC.

At the broadest scales in Fig. 8, the EnKF full-system experiment is the odd one out, producing wind increments everywhere. The other three experiments give significant increments in three areas on this

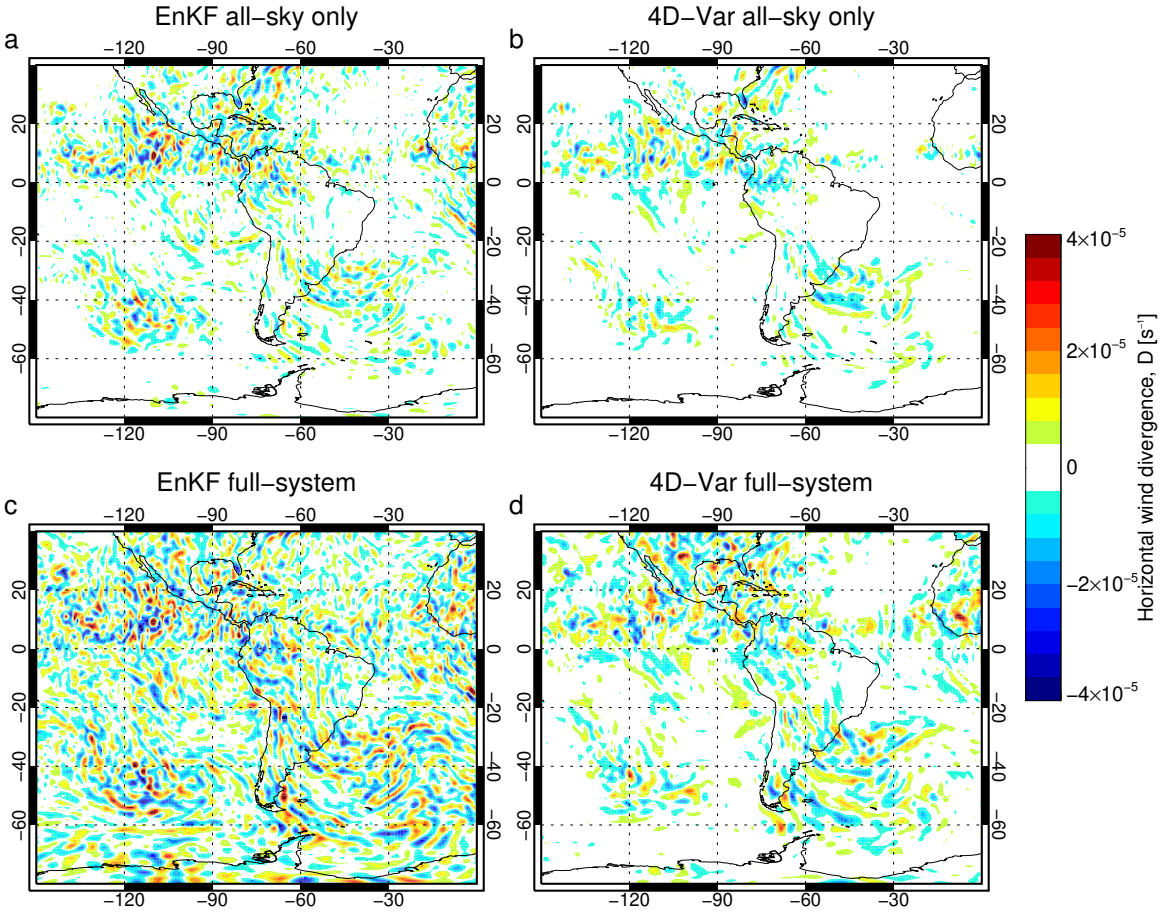

Figure 8: Increments in horizontal wind divergence on model level 74 (approximately 200 hPa) at 00 UTC 15th August 2015. 'All-sky only' are experiments assimilating only all-sky data; 'Full-system' are experiments assimilating the full global observing system. Model levels are terrain-following and adapt to the local pressure field, so level 74 is at 197 hPa if the surface is at standard temperature and pressure. The full model resolution has been truncated to 1 degree before plotting. The initial conditions in panels a-c are from the full-system EnKF experiment and in panel d from the full-system 4D-Var experiment.

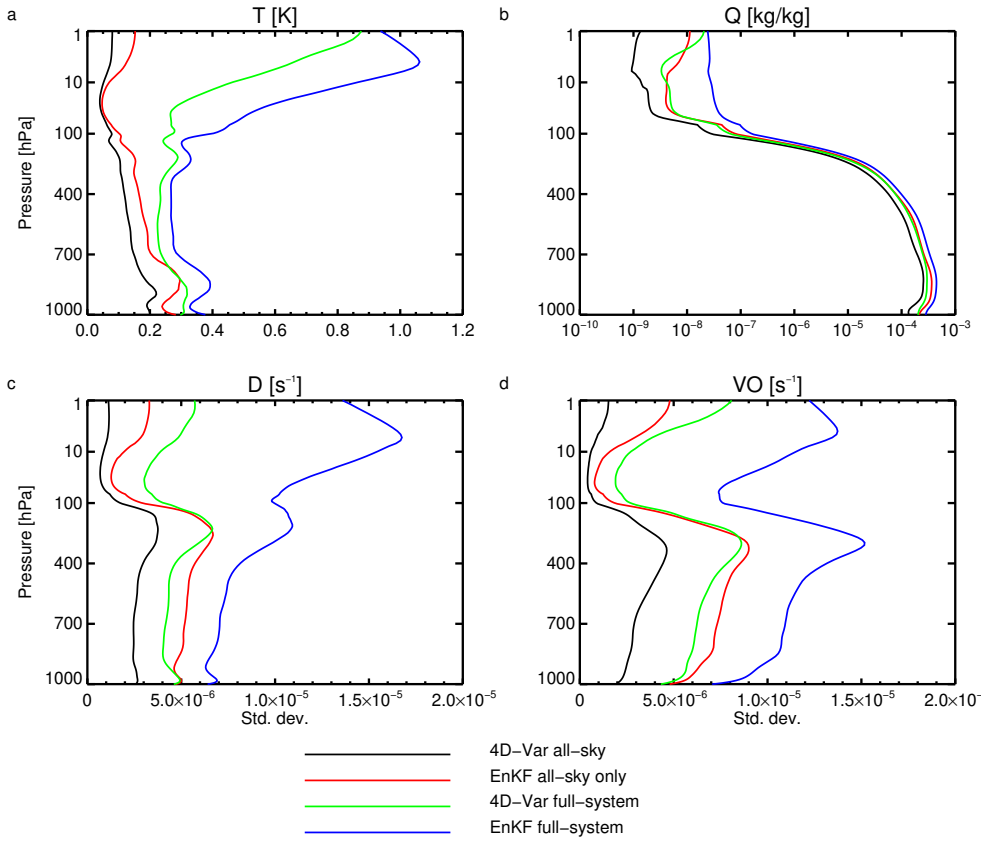

Figure 9: Standard deviations of data assimilation increments in different experiments at 00 UTC on 15th August 2015. Correlations are 'smoothed' in the vertical (see text). T stands for temperature, Q specific humidity, D horizontal wind divergence, VO horizontal vorticity. The pressure scale is linear in the troposphere (below 100hPa) and logarithmic in the stratosphere.

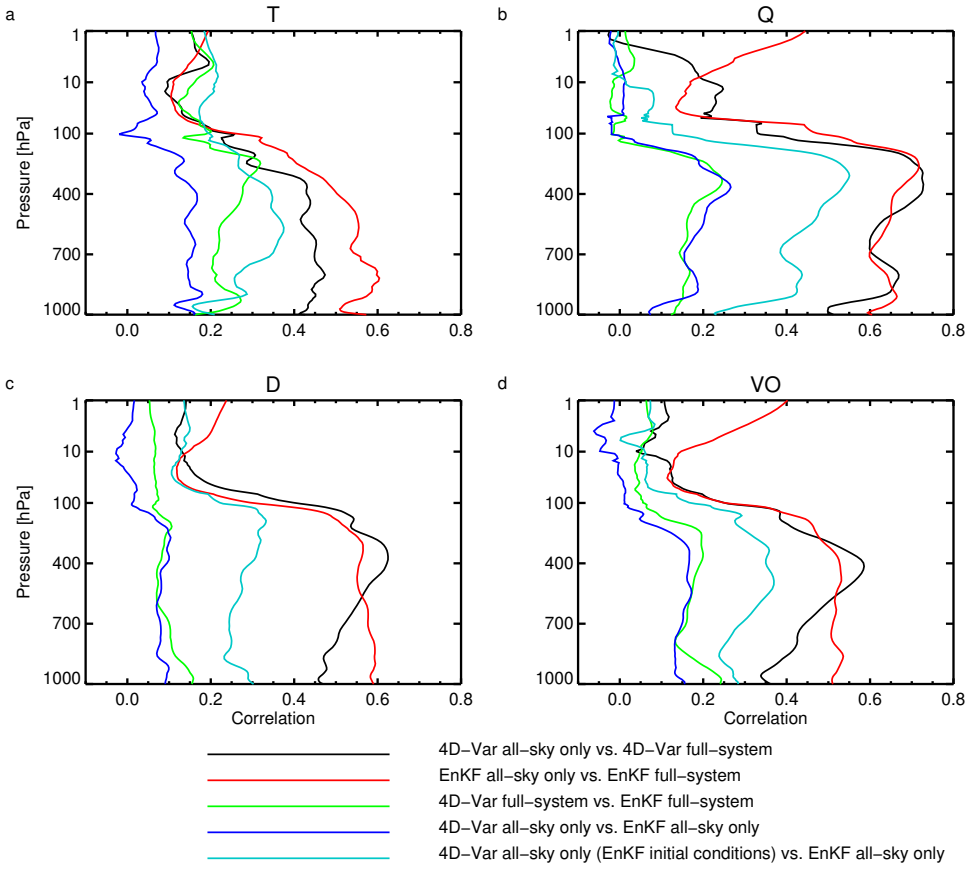

Figure 10: Correlations in data assimilation increments between different experiments at 00 UTC on 15th August 2015. Other details as Fig. 9

figure, in the Intertropical-Convergence Zone (ITCZ) and two areas of active midlatitude weather in the SE Pacific and S Atlantic. Both the all-sky and full-system EnKF experiments seem to generate increments of larger magnitude and smaller spatial scale than the equivalent 4D-Var experiments. To generalise this, Fig. 9 shows the standard deviation of global increments as a function of pressure in the vertical. To smooth the figure, a sliding window of 21 vertical model levels has been used in the stratosphere and 5 model levels in the troposphere, with a smooth transition between the two at around 100hPa. Standard deviations are computed from a sample encompassing all the selected model levels and global increments on the regular 1-degree lat-lon grid. This confirms the EnKF generates increments of much higher standard deviations than 4D-Var, particularly in the stratosphere, not just in the wind variables but in temperature as well. The discrepancy is largest around 10hPa in divergence, where increment standard deviations are about 15×10^{-6} s^{-1} in the full-system EnKF and just 4×10^{-6} s^{-1} in the full-system 4D-Var. Stratospheric increments are very much smaller in the all-sky only experiments. Down into the troposphere, in temperature and specific humidity, the 4D-Var and EnKF standard deviations are relatively similar, but in divergence and vorticity, EnKF increments are still around 60% larger than in 4D-Var. Fig. 8 also shows many similarities between the wind increments in EnKF and 4D-Var, particularly between the all-sky-only experiments. The similarities are strongest in the areas of active weather, for example over the Atlantic just off the coast of Brazil. These wind increments are also similar to those in the full-system experiments, both with EnKF and 4D-Var.

Comparisons to observations in subsection 4.5 show our EnKF implementation looks worse in the stratosphere. An issue of excessive gravity wave increments was already recognised and partially addressed using a divergence adjustment technique (Hamrud et al., 2015; Bonavita et al., 2015). However this aims to stabilise the surface pressure tendency, which is dominated by the mass in the mid and lower troposphere. Hence it does not much constrain the generation of gravity wave increments in the upper troposphere and stratosphere. The excessive stratospheric standard deviations are not seen in the all-sky-only EnKF experiment, so the problem must come from the assimilation of other observation types, likely those directly sensitive to the stratosphere. In 4D-Var, the generation of stratospheric gravity wave increments is suppressed in two main ways. First the mass-wind balance of the analysis increments is progressively deactivated starting upwards from 50 hPa, in such a way that the analysis becomes effectively univariate above 20 hPa. Secondly, specific parts of the balance operators connecting the control space of the analysis to the state space of the model are zeroed-out or regularised because the underlying sample statistics appear noisy or unphysical. The net result of this treatment of the 4D-Var balance operator is a reduction of the divergent wind increment.

Many of the increments in the mid and upper-troposphere that are common between 4D-Var and EnKF in Fig. 8 are wave-like patterns with wavelengths of order 500 km. We speculate that in the midlatitudes these are associated with mesoscale gravity waves that are often generated by non-geostropic adjustment of Rossby waves (e.g. O'Sullivan and Dunkerton, 1995; Pavelin et al., 2001, in these works gravity waves are more precisely called inertia-gravity waves, IGW). The data assimilation system must correct both the larger (e.g. Rossby) scales but also the gravity wave structures on the mesoscale. These structures (and similar wave-like structures in the tropics) are of genuine observational origin as they are, even in 4D-Var, one of the most obvious features of all-sky background departures in water vapour channels sensitive to the mid and upper-troposphere (see the presentation of A. Geer at ITSC XXI, available from `https://cimss.ssec.wisc.edu/itwg/itsc/itsc21/program/index.html`). The semi-implicit semi-Lagrangian advection in the IFS heavily damps gravity waves and cannot correctly model their phase speeds (Simmons and Temperton, 1997; Hamrud et al., 2015). Therefore every analysis cycle, the analysis system has to re-construct mesoscale gravity waves that have become damped and mislocated. 4D-Var allows this in the active parts of the upper-troposphere and (based on the success of 4D-Var) this likely adds important information to the initial conditions that helps improve the long-range forecasts.

But EnKF also allows this in the quiet areas of the upper-troposphere and the stratosphere, which may degrade forecasts, and certainly degrades the early-range fits to observations (see later). Whether the gravity waves generated in these areas are physically realistic or whether they are partially the result of noisy correlations in the EnKF would need further research.

To further examine the similarity of increments between EnKF and 4D-Var, Fig. 10 shows the global correlations between the different experiments, as a function of vertical level. There are relatively good correlations between increments in the all-sky only EnKF and 4D-Var experiments that share the same initial conditions (the light blue line). These are around 0.25 to 0.35 through most of the troposphere for temperature, divergence and vorticity, and as high as 0.4 to 0.5 for specific humidity. This reflects the agreement seen visually between panels a and b in Fig. 8. If the all-sky only 4D-Var experiment uses the initial conditions from the full-system 4D-Var experiment, correlations drop to around 0.1 to 0.2. The biggest correlations are between all-sky-only and full-system, whether in the EnKF or 4D-Var, which range from 0.4 to 0.6 for all four variables throughout the troposphere. Similar correlations were shown by Geer et al. (2014) in the 4D-Var context in the mid and upper-troposphere, but their correlations were smaller in the lower troposphere because their study did not include microwave imager channels, which are more sensitive to the lower levels (e.g. Fig. 4, bottom row). Overall it seems that all-sky observations are providing similar information in EnKF and 4D-Var (even if EnKF seems to generate slightly smaller spatial scales and larger increments) and both appear to be using a form of wind tracing to improve the wind fields. This should not be a surprise given the correlations shown in Fig. 4; also for example Allen et al. (2015) have demonstrated in a more theoretical context how wind-tracing can be achieved in an EnKF.

4.4 Comparing the impact of all-sky observations in LETKF and 4D-Var

As will be seen in the next subsection, the absolute quality of the EnKF configuration remains lower than that of 4D-Var. However, it is interesting to contrast the relative impact of all-sky observations in the two systems. This is shown in Fig. 11, in each case on top of the high-quality baseline system containing all other observations. At the broadest level, the impact of all-sky assimilation is equally beneficial in the EnKF and 4D-Var, giving around a 2% – 4% improvement. Here we concentrate on forecast scores in the medium range (beyond day 3) and particularly in geopotential height at 500 hPa and wind vector at 850 hPa. For these variables, and at these ranges, synoptic errors grow rapidly, so that the choice of verifying analysis is not too important and these results are robust. Despite the experiments being performed at lower resolution and on a 6 h assimilation window, this impact is also broadly comparable to the results for the high resolution, T1279, 12h operational cycle 41r1 4D-Var configuration (Geer et al., 2017).

Similar scores were also generated using each experiment's own-analysis as the reference (not shown). For these there, in the short-range (up to day 3), in the tropics, and particularly in relative humidity (RH), there were large apparent differences between 4D-Var and EnKF. For example the impact of all-sky in 4D-Var went from a 20% 'improvement' in tropical 850 hPa RH to what is likely an equally unreliable 35% 'degradation'. In contrast EnKF apparently still improved forecast quality by around 10%. However, in these variables and ranges, there are correlations between analysis and forecast errors as well as systematic errors that can be an important part of the RMS error. All-sky assimilation of microwave imagers in 4D-Var does a lot of work to fit the observed patterns of oceanic boundary layer moisture, cloud and wind (Geer and Bauer, 2010). This adds substantial perturbations to the analysis that are diffused away over the next few days, but they can show up as an apparent degradation in own-analysis forecast scores. It is very interesting, but not currently explained, that EnKF does not suffer

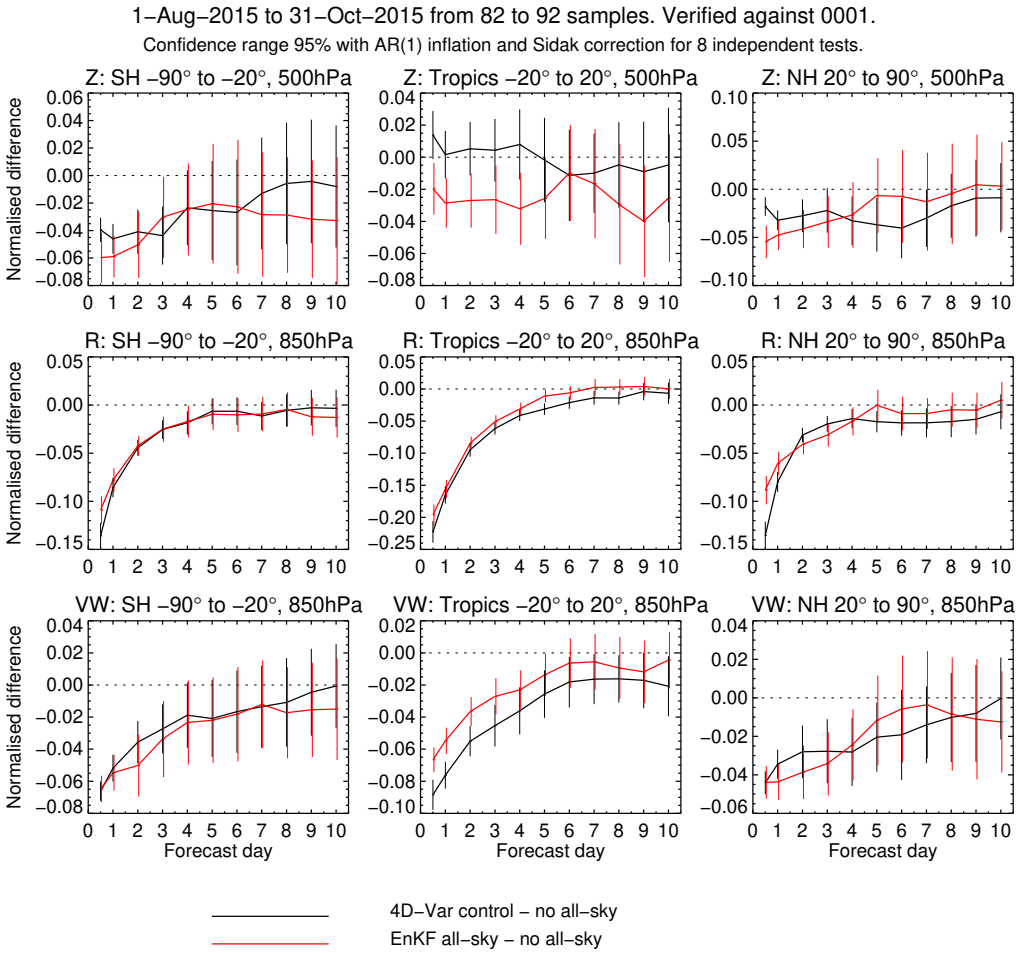

Figure 11: Impact of all-sky assimilation in 4D-Var and LETKF configurations, using the spread-multiple error model and 50 members in the case of the LETKF run. Results use the operational 4D-Var analysis as the verification reference in each case, and show the change in the RMS errors between an experiment and a control, normalised by the errors in the control.

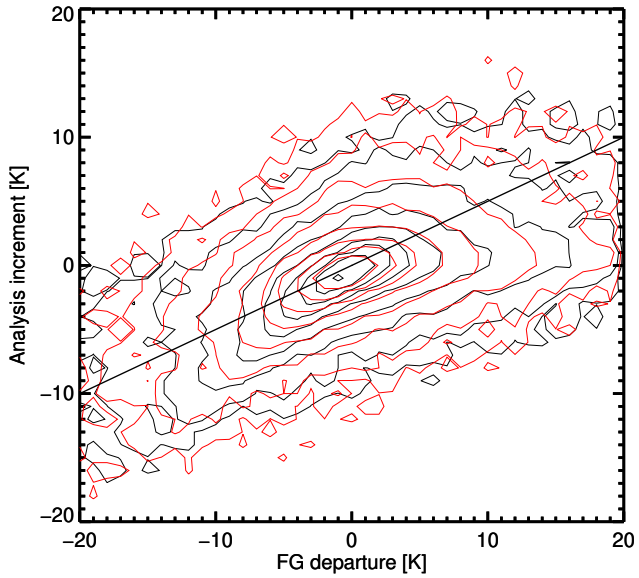

Figure 12: Joint histograms of background departures and increments in SSMIS F-17 channel 19v from 4D-Var (black) and EnKF AM2 (red) configurations. Based on all assimilated data from 1-9 September 2015. The 1:2 line is overplotted. EnKF increments are taken from the control member, i.e. they are the ensemble mean analysis minus the background forecast from the previous ensemble mean analysis. Contours are logarithmically spaced, starting with 3, 10, 32, 100, 316 ... observations per 1 K by 1 K bin.

this effect. The background fits to other observations (not shown, but similar to Fig. 6) are significantly and consistently improved by all-sky assimilation in 4D-Var and EnKF, confirming that analysis-based scores (even using the operational analysis as a reference) likely do not show the true magnitude or sign of the improvement or degradation. Hence all the apparent impacts in analysis-based forecast scores in relative humidity and in all variables before day 3 should be treated with extreme caution.

Figure 12 compares the joint histograms between background departures and increments from the 4D-Var and EnKF experiments for SSMIS channel 19v. If the analysis were to exactly fit the observations, then all points would be on the 1:1 line. The gradient is in practice smaller because the analysis is a balance between prior and observational information, and there is scatter because the analysis has to fit many other observations, not just the all-sky data. For the same departures, EnKF and 4D-var produce surprisingly similar increments in observation space. Hence EnKF and 4D-Var increments do not seem to differ much at the observation locations. This is consistent with the increments in model space, which are partly correlated between 4D-Var and EnKF. The strong similarity of the observation-space increments in 4D-Var and EnKF also further justifies a need for medium-range forecast verification statistics to reliably identify performance differences between the different systems.

Previous studies of the joint histogram between background departures and increments (Fig. 12) have shown that the ECMWF 4D-Var system is more able to dry (i.e. to remove cloud and precipitation) than to moisten (i.e. to create cloud and precipitation) (e.g. Geer et al., 2010). Negative departures, which for channel 19v indicate the model should dry, here generate increments that are clustered approximately around the 1:2 line. Positive departures, which require a moistening, are associated with much smaller increments that are on average closer to the 0 line. Previously the 4D-Var linearised moist physics was suspected as a possible source of the asymmetry. If the EnKF system has similar behaviour, the relative inability of data assimilation systems to moisten in response to all-sky observations may come

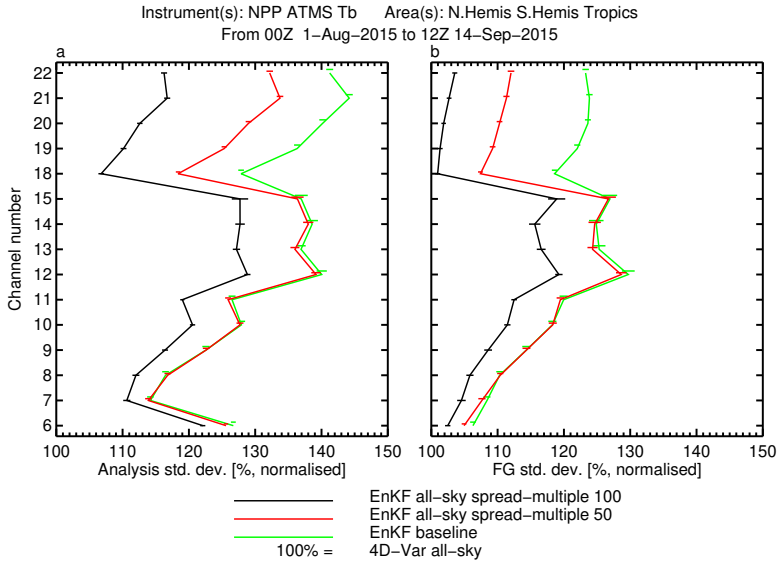

Figure 13: Standard deviation of normalised analysis (left panel) and first guess (right panel) ATMS brightness temperature departures in some of the EnKF experiments, normalised with respect to the departures in the all-sky 4D-Var experiment. Values above 100 indicate worse performance with respect to the 4D-Var.

from something more fundamental, such as the boundedness of the problem (e.g. Geer and Bauer, 2011; Posselt and Bishop, 2018).

4.5 Quality of LETKF compared to 4D-Var

Comparing the EnKF and 4D-Var experiments that include all-sky assimilation confirms the findings of Bonavita et al. (2015) that results from the EnKF are not as good as from 4D-Var. Figures 13 and 14 show the analysis and first guess fit of ATMS and radiosonde temperature observations of the previously-examined 50 member EnKF experiments with and without all-sky assimilation, as well as an additional 100 member experiment that will be discussed shortly. All-sky assimilation brings improvements but this is not nearly enough to close the gap. The analysis error standard deviations of the 50 member EnKF experiments are significantly larger than the 4D-Var fit, by around 20 to 30% in the stratosphere (e.g. ATMS channel 12) but reducing to less than 5% of 4D-Var in the troposphere (e.g. ATMS channel 6). Similar results are visible with other major observing systems as the reference. The most obvious impact of all-sky assimilation is on the humidity-sensitive channels of ATMS, where FG errors are reduced from around 20% to 10% worse than 4D-Var. The impact on temperature channels and radiosonde fits is much smaller.

Forecast performance in the medium-range is consistent with the observation departure statistics. Figure 15 compares forecast skill of the EnKF experiments to that of 4D-Var for the 500 hPa geopotential forecast. Even with all-sky assimilation, the EnKF has a consistent degradation of performance of 5 to 10% through much of the forecast range. In the stratosphere the gap in forecast performance is even larger (20 to 50%, not shown), consistent with the larger first guess standard deviations seen in Figure 14. The impact of adding all-sky assimilation is a step along the way to matching 4D-Var, apparently improving forecasts by around 5% at day 2. However, larger improvements can be made by changing the ensemble size, as already seen by Hamrud et al. (2015). This is confirmed by increasing the ensemble size from 50 to 100 in the current work, leaving all other parameters unchanged from the 50-member

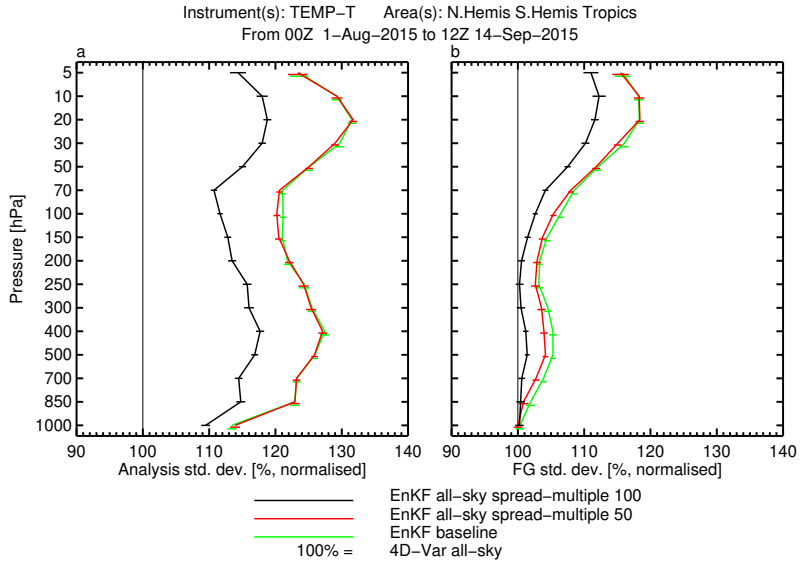

Figure 14: Standard deviation of normalised analysis (left panel) and first guess (right panel) radiosonde temperature departures, normalised with respect to the departures in the all-sky 4D-Var experiment.

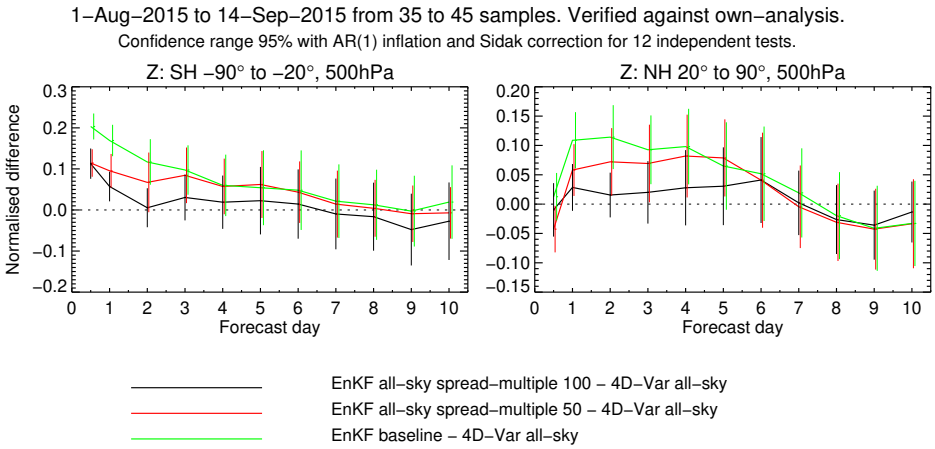

Figure 15: Normalised change in RMS error in the southern extra-tropics (left panel) and northern extra-tropics (right panel) of the 500 hPa geopotential forecast for EnKF experiments with respect to the all-sky 4D-Var experiment. Values above the zero line indicate worse performance with respect to 4D-Var. Error bars indicate 95% confidence levels. Verification against own analysis.

experiment with the spread-multiple error model. The results in Figs. 13, 14 and 15 show a significant improvement in analysis and forecast skill with the increase in ensemble size.

We have seen that increments are larger in the EnKF than in 4D-Var (Fig. 9) but the analysis and forecast fits to observations are worse (Fig. 14). It must be emphasized that in an ideal world more effort could have been invested in optimizing the performance of the EnKF. It might be possible to improve the analysis fit to observations by tuning the relative observation errors (since the 4D-Var defaults were used) or the implied background errors (through changes to localisation and inflation parameters). A possible explanation may be the excessive increments resembling gravity waves, particularly in the upper troposphere and stratosphere (Fig. 9). Increasing the ensemble size may partially address this by reducing spurious correlations and by reducing spurious variability in the background error standard deviations. There might be cheaper ways to smooth and/or filter either the implied covariances or the increments. However, with current approaches and in a pure EnKF context, this work further confirms that ensemble sizes of at least O(100) are needed to obtain good results with global NWP systems using contemporary observing systems.

5 Conclusions

We have presented a comprehensive examination of all-sky radiance assimilation within an ensemble Kalman filter. The advantages of this study over previous work are that (i) it uses a framework that is close to operational quality for global weather forecasting; (ii) it assimilates a full suite of all-sky microwave observations from 8 sensors, globally over ocean, sea-ice and land scenes; (iii) most experiments have been run for 3 months duration to allow statistical significance testing. We sought to understand whether the assimilation of all-sky observations might challenge the linear and Gaussian restrictions of the pure ensemble approach. We also wanted to understand whether ensemble systems can make practical use of the tracing effect, which enables the extraction of indirectly observed information, particularly winds, from all-sky observations sensitive mainly to humidity, cloud and precipitation. This mechanism is thought to provide much of the benefit of all-sky observations to medium-range forecast scores in 4D-Var (Geer et al., 2017, 2018). Further, the ability of incremental 4D-Var to make use of non-linear and non-Gaussian observations is also considered a major advantage (Bauer et al., 2010; Bonavita et al., 2018).

We compared versions of the IFS using the EnKF and a 4D-Var using static background error covariances and a 6h assimilation window (this is a cleaner comparison than against the hybrid background errors and 12h assimilation window used in operations). The increments generated from all-sky observations had similar patterns in the EnKF and in 4D-Var with correlations between them of around 0.3 in temperature, divergence and vorticity, and around 0.4 to 0.5 in specific humidity (when identical backgrounds were used). Further, the impact of adding all-sky observations on medium range forecast errors was around 2% – 4% in both EnKF and 4D-Var (consistent with results at the full operational 4D-Var resolution, Geer et al., 2017). The generation of wind increments confirms the expectation from studies, with more limited assimilation systems, that an EnKF is also perfectly capable of generating wind tracing from constituents like water vapour (e.g. Allen et al., 2015) and also from cloud and precipitation observations (e.g. Lien et al., 2016).

The ensemble correlations in the EnKF reveal the unique information content brought by all-sky microwave observations, particularly their strong indirect sensitivity to the wind field. Typical temperature-sounding microwave channels assimilated in clear-sky conditions have broad correlations in the vertical (e.g. the whole troposphere) and horizontal (out to 500 km) and relatively weak but distant wind cor-

relations. In contrast the water-vapour sounding and imaging channels that are assimilated in all-sky conditions have sensitivities that are more localised (to perhaps half the troposphere, and to 200 km horizontally) but they have stronger sensitivity to wind. Although there are known advantages of incremental 4D-Var for handling strongly nonlinear observations such as all-sky microwave radiances, the linearity inherent in the standard EnKF does not appear to block doing successful all-sky assimilation in a high quality EnKF. Further, the relatively compact ensemble correlations associated with all-sky microwave observations is likely to make them easier to implements in an EnKF framework. Our results back up the generally positive results in the literature demonstrating all-sky assimilation mainly in shorter EnKF experiments with less extensive observing systems. It must be noted, however, that cycling the assimilation every 6 hours (instead of the customary 12 hours used at ECMWF) and using a 4D-Var test configuration with much closer matching of outer-inner loop resolution than that used in operations, has likely reduced the impact of nonlinearities and non-Gaussian effects in our experiments. More generally, efficiently dealing with nonlinearities and the resulting non-Gaussian effects in operational NWP can be achieved through: more frequent analysis updates; repeated re-linearisations in the analysis algorithm (outer loop mechanism in 4D-Var; iterated EnKF, eg (Evensen, 2018) and references therein); a combination of the two. The algorithmic solution of choice will likely be dictated not simply by the absolute accuracy but also by other aspects which are important in an operational context (eg, computational efficiency, time to solution, scalability on available computing architectures).

The similarity of results in the EnKF also extends to a problem that has affected the IFS 4D-Var for many years: it is difficult to moisten the analysis, particularly to create cloud and precipitation when required (e.g. Geer et al., 2010). In the EnKF this might be explained by the zero-gradient problem or in 4D-Var it might be difficulties propagating moistening increments through the TL and adjoint models, or by the suppression of moistening increments in the humidity control variable. A more general explanation might come from the boundedness of the moist variables and the observations we have of them (e.g. Hólm et al., 2002; Geer and Bauer, 2011; Bishop, 2019). Humidity, cloud and precipitation all have a lower bound of zero and an upper bound set by 100% relative humidity in one case and by the processes of condensation and precipitation in the other. Whatever the explanation, there remains a general problem in creating (rather than destroying) cloud and precipitation in current operational data assimilation methods.

A substantial effort was needed to get the EnKF working with all-sky observations. We obtained around 1% better results (in terms of short-range forecast quality measured against other observations) by reducing the vertical and horizontal localisation scales compared to those used for clear-sky radiances and conventional observations (from 2.0 to 1.6 times the scale height, and from 2000 km to 1000 km). The choice of error model had an even larger influence on the short-range forecast quality. In addition to testing the commonly used 'symmetric' observation error model, which degraded tropospheric temperature forecasts at short range in our experiments, we proposed two new possible observation error models for use in an ensemble Kalman filter framework (note that observation error correlations are not yet considered, only variances). First is a 'nonlinearity' approach based on the consistency, in observation space, between the ensemble mean and the unperturbed control forecast. The second new model assigns observation error as a multiple of the prior spread in observation space, assuming that the major error in cloud and precipitation-affected observations is an error of representation and/or model error with similar characteristics to the background error. The key distinctions of these new error models compared to the symmetric approach are that (i) they do not rely on the observation, but rather on the background ensemble, to inflate the observation error in the presence of cloud; and (ii) they take account of the presence of cloud in any of the ensemble members, rather than just in the ensemble mean or control forecast.

The 'spread multiple' model worked best in these tests, for example allowing all-sky observations to make a 1.5% improvement in fit to ATMS channel 6 observations (a channel with mid tropospheric tem-

perature sensitivity) as compared to 1% improvement in the next-best model, the nonlinearity approach, and a 0.2% degradation with the symmetric error model. However we did not have a chance to explore all the possible refinements to these models. For example the 'nonlinearity' error model was tested with its tunable parameter $\alpha = 1.0$, which give it the smallest observation errors in cloudy situations compared to the other two models, so it may have been overfitting cloudy observations in some situations. The most successful error model, the spread multiple, generated total errors that were significantly larger than the standard deviation of the ensemble mean first guess departures. This would likely have compensated for sub-optimalities in the data assimilation, either from nonlinearity or non-Gaussianity affecting the EnKF, or the effect of observation error correlations that have not been fully accounted for (common to all configurations here). Further work is needed to see whether differences between the models are fundamental or just down to tuning. Further, the weighting parameters should not be globally constant, but should vary depending on the channel.

There remains great interest in seeing whether ensemble data assimilation (concentrating here specifically to the use of ensemble correlations to propagate background errors through the assimilation time window) can start to challenge the quality of global forecasts made using 4D-Var (which uses TL and adjoint models to propagate background errors). Howver, adding all-sky assimilation is beneficial but does not change the conclusions of Hamrud et al. (2015). In the troposphere, the 50-member EnKF is still around 5% – 10% worse than 4D-Var. A further and bigger improvement in the quality of the EnKF, of around 5%, comes from increasing the ensemble size from 50 to 100 members, consistent with results from many other studies. This brought the EnKF to within 2% to 5% of the 4D-Var performance in the troposphere. Here we have compared EnKF to 4D-Var with a static (i.e. climatological) background error covariance. In practice, operational ensemble and 4D-Var systems use hybrid configurations, which do better than the individual components (e.g. Bonavita et al., 2015). It would be hard to extrapolate from our results to say whether an improved EnKF could be part of a system that would challenge the current IFS hybrid 4D-Var. Lorenc and Jardak (2018) have shown that the use of TL and adjoint models in variational assimilation continues to outperform the use of ensemble correlations.

A remaining issue with the IFS EnKF is poor performance in the stratosphere (Hamrud et al., 2015; Bonavita et al., 2015). This is not directly related to all-sky assimilation. However the fitting of observed mesoscale (e.g. order 500 km) gravity waves appears a necessary part of both the EnKF and 4D-Var analysis in many places, for example around active weather systems in the midlatitudes in the upper troposphere. When assimilating just all-sky observations, these gravity wave increments are similar in EnKF and 4D-Var, and are thought to be beneficial. However, other gravity waves generated by the EnKF using the full observing system appear spurious by comparison to 4D-Var, particularly in quieter areas of the upper-troposphere and throughout the stratosphere. These gravity waves may be responsible for degrading the background fits of the EnKF to temperature-sensitive observations, which are 10% to 30% worse than 4D-Var in the stratosphere. The EnKF increments were also generally larger (by up to 60% in the troposphere) and were made on slightly smaller spatial scales than in 4D-Var. Adding ensemble members mitigated this but it might be an expensive way of addressing the problem. The stratospheric observing system is dominated by nadir sounding microwave observations which sample the atmosphere over deep vertical layers. Vertical covariance localisation probably limits the amount of information that can be extracted from these observations, although new techniques have been proposed to address this (e.g. Mitchell et al., 2018; Shlyaeva et al., 2019). However there are deep stratospheric error modes in the IFS that are connected to slowly evolving model biases, and hence cannot be effectively represented by the "errors of the day" of the EnKF (or any other online ensemble data assimilation system). Normal mode initialisation could also be helpful (e.g. Allen et al., 2015). More generally, our assessment of the quality of the EnKF can never be definitive as (given sufficient time) there may be many ways to move forward.

A final factor is cost. Even if a system is affordable for operations, the development of high-quality operational forecast systems relies on plentiful experimentation. It is not just a matter of 'inspiration' – major developments like the move to 4D-Var, the introduction of hybrid background error covariances, or the ability to use cloudy and precipitating observations – but also on 'perspiration', the everyday grind of finding useful incremental improvements by running thousands of exploratory experiments over many years. We have not been able to fully explore the possible space of improvements that could make the EnKF more competitive because our EnKF configuration is approximately 3 times more costly than 4D-Var with static error covariances. An additional practical issue on the ECMWF supercomputer was the use of a 6 h assimilation window exposed our experiments to twice as much queueing time as a 12 h window assimilation experiment, compounding the problem. Experimentation would have been harder still had we used a 100 member ensemble throughout. Hence an argument against further development of the EnKF version of the IFS (or, more generally, any pure ensemble-based analysis algorithm) would be the reduced amount of routine exploratory experimentation that could be afforded to scientists seeking to improve the system.

Acknowledgements

The authors thank Niels Bormann, Stephen English and Andy Brown for internal reviews of the manuscript, Elias Hólm for helping understand the ensemble error budget, and Ed Pavelin for discussions on gravity waves.

References

Allen, D., K. Hoppel, and D. Kuhl (2015). Wind extraction potential from ensemble Kalman filter assimilation of stratospheric ozone using a global shallow water model. *Atmos. Chem. Phys 15*, 5835–5850.

Baordo, F. and A. J. Geer (2016). Assimilation of SSMIS humidity-sounding channels in all-sky conditions over land using a dynamic emissivity retrieval. *Quart. J. Roy. Meteorol. Soc. 142*, 2854–2866.

Bauer, P., A. J. Geer, P. Lopez, and D. Salmond (2010). Direct 4D-Var assimilation of all-sky radiances: Part I. Implementation. *Quart. J. Roy. Meteorol. Soc. 136*, 1868–1885.

Bishop, C. H. (2019). Data assimilation strategies for state-dependent observation error variances. *Quart. J. Roy. Meteorol. Soc. 145*(718), 217–227.

Bonavita, M., M. Hamrud, and L. Isaksen (2015). EnKF and hybrid gain ensemble data assimilation. Part II: EnKF and hybrid gain results. *Mon. Weath. Rev. 143*(12), 4865–4882.

Bonavita, M., L. Isaksen, E. Hólm, and M. Fisher (2016). The evolution of the ECMWF hybrid data assimilation system. *Quart. J. Roy. Meteorol. Soc. 142*, 287–303.

Bonavita, M., P. Lean, and E. Holm (2018). Nonlinear effects in 4D-Var. *Nonlin. Proc. Geophys. 25*(3), 713–729.

Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally (2016). Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteorl. Soc. 142*(697), 1767–1780.

Bormann, N., A. Geer, and P. Bauer (2011). Estimates of observation error characteristics in clear and cloudy regions for microwave imager radiances from NWP. *Quart. J. Roy. Meteorol. Soc. 137*, 2014–2023.

Bormann, N., H. Lawrence, and J. Farnan (2019). Global observing system experiments in the ECMWF assimilation system. Tech. Memo. 839, ECMWF, Reading, UK.

Buehner, M., R. McTaggart-Cowan, A. Beaulne, C. Charette, L. Garand, S. Heilliette, E. Lapalme, S. Laroche, S. R. Macpherson, J. Morneau, et al. (2015). Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part I: The global system. *Mon. Wea. Rev. 143*(7), 2532–2559.

Cardinali, C. (2009). Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteorol. Soc. 135*(638), 239–250.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteorol. Soc. 120*, 1367–1387.

Dee, D. P. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weath. Rev. 123*(4), 1128–1145.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteorol. Soc. 131*, 3385–3396.

Errico, R. M., P. Bauer, and J.-F. Mahfouf (2007). Issues regarding the assimilation of cloud and precipitation data. *J. Atmos. Sci. 64*, 3785 – 3798.

Evensen, G. (2018). Analysis of iterative ensemble smoothers for solving inverse problems. *Comput. Geosci. 22*, 885–908.

Fletcher, S. and A. Jones (2014). Multiplicative and additive incremental variational data assimilation for mixed lognormalgaussian errors. *Mon. Wea. Rev. 142*, 2521–2544.

Geer, A. J. (2016). Significance of changes in forecast scores. *Tellus A 68*, 30229, http://dx.doi.org/10.3402/tellusa.v68.30229.

Geer, A. J. (2019). Correlated observation error models for assimilating all-sky infrared radiances. *Atmos. Meas. Tech. 12*, 3629–3657.

Geer, A. J., F. Baordo, N. Bormann, and S. English (2014). All-sky assimilation of microwave humidity sounders. Tech. Memo. 741, ECMWF, Reading, UK.

Geer, A. J., F. Baordo, N. Bormann, S. English, M. Kazumori, H. Lawrence, P. Lean, K. Lonitz, and C. Lupu (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteorol. Soc. 143*, 3189–3206.

Geer, A. J. and P. Bauer (2010). Enhanced use of all-sky microwave observations sensitive to water vapour, cloud and precipitation. Tech. Memo. 620, ECMWF, Reading, UK.

Geer, A. J. and P. Bauer (2011). Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteorol. Soc. 137*, 2024–2037.

Geer, A. J., P. Bauer, and P. Lopez (2010). Direct 4D-Var assimilation of all-sky radiances: Part II. Assessment. *Quart. J. Roy. Meteorol. Soc. 136*, 1886–1905.

Geer, A. J., K. Lonitz, P. Weston, M. Kazumori, K. Okamoto, Y. Zhu, E. H. Liu, A. Collard, W. Bell, S. Migliorini, P. Chambon, N. Fourrié, M.-J. Kim, C. Köpken-Watts, and C. Schraff (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quart. J. Roy. Meteorol. Soc. 144*, 1191–1217.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt (2011). Balance and ensemble Kalman filter localization techniques. *Monthly Weather Review 139*, 511–522.

Gürol, S., A. Weaver, A. Piacentini, H. Arango, and S. Gratton (2014). B–preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Quart. J. Roy. Meteorol. Soc. 140*, 539–558.

Hamrud, M., M. Bonavita, and L. Isaksen (2015). EnKF and hybrid gain ensemble data assimilation. Part I: EnKF implementation. *Mon. Weath. Rev. 143*(12), 4847–4864.

Harnisch, F., M. Weissmann, and Á. Periáñez (2016). Error model for the assimilation of cloud-affected infrared satellite observations in an ensemble data assimilation system. *Quart. J. Roy. Meteorol. Soc. 142*(697), 1797–1808.

Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J.-F. Mahfouf, A. Simmons, and J.-N. Thepaut (2002). Assimilation and modelling of the hydrological cycle: ECMWF's status and plans. Tech. Memo. 383, ECMWF, Reading, UK.

Honda, T., T. Miyoshi, G.-Y. Lien, S. Nishizawa, R. Yoshida, S. A. Adachi, K. Terasaki, K. Okamoto, H. Tomita, and K. Bessho (2018). Assimilating all-sky Himawari-8 satellite infrared radiances: A case of Typhoon Soudelor (2015). *Mon. Weath. Rev. 146*(1), 213–229.

Houtekamer, P. and F. Zhang (2016). Review of the ensemble Kalman filter for atmospheric data assimilation. *Mon. Weath. Rev. 144*(12), 4489–4532.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Mon. Weath. Rev. 133*(3), 604–620.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena 230*(1), 112–126.

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud (2010). Ensemble of data assimilations at ECMWF. Tech. Memo. 636, ECMWF, Reading, UK.

Kazumori, M., A. J. Geer, and S. J. English (2016). Effects of all-sky assimilation of GCOM-W/AMSR2 radiances in the ECMWF numerical weather prediction system. *Quart. J. Roy. Meteorol. Soc. 142*, 721–737.

Kleist, D. T. and K. Ide (2015). An OSSE-based evaluation of hybrid variational–ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Mon. Wea. Rev. 143*(2), 452–470.

Langland, R. H. and N. L. Baker (2004). Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A 56*(3), 189–201.

Lawrence, H., N. Bormann, A. Geer, Q. Lu, and S. English (2018). Evaluation and assimilation of the microwave sounder MWHS-2 onboard FY-3C in the ECMWF numerical weather prediction system. *IEEE Trans. Geosci. Rem. Sens. 56*, 3333 – 3349.

Lean, P., A. Geer, and K. Lonitz (2017). Assimilation of Global Precipitation Mission (GPM) Microwave Imager (GMI) in all-sky conditions. Tech. Memo. 799, ECMWF, Reading, UK.

Lien, G.-Y., T. Miyoshi, and E. Kalnay (2016). Assimilation of TRMM multisatellite precipitation analysis with a low-resolution NCEP global forecast system. *Mon. Weath. Rev. 144*(2), 643–661.

Lorenc, A. C. and M. Jardak (2018). A comparison of hybrid variational data assimilation methods for global NWP. *Quart. J. Roy. Meteorol. Soc. 144*(717), 2748–2760.

Migliorini, S. and B. Candy (2019). All-sky satellite data assimilation of microwave temperature sounding channels at the Met Office. *Quart. J. Roy. Meteorol. Soc. 145*(719), 867–883.

Minamide, M. and F. Zhang (2017). Adaptive observation error inflation for assimilating all-sky satellite radiance. *Mon. Weath. Rev. 145*, 1063–1081.

Minamide, M. and F. Zhang (2019). An adaptive background error inflation method for assimilating all-sky radiances. *Quart. J. Roy. Meteorol. Soc. 145*(719), 805–823.

Mitchell, H., P. Houtekamer, and S. Heilliette (2018). Impact of AMSU-A radiances in a column ensemble Kalman filter. *Mon. Wea. Rev. 146*, 3949–3976.

Okamoto, K., A. P. McNally, and W. Bell (2014). Progress towards the assimilation of all-sky infrared radiances: an evaluation of cloud effects. *Quart. J. Roy. Meteorol. Soc. 140*(682), 1603–1614.

Okamoto, K., Y. Sawada, and M. Kunii (2019). Comparison of assimilating all-sky and clear-sky infrared radiances from Himawari-8 in a mesoscale system. *Quart. J. Roy. Meteorol. Soc. 145*, 745–766.

O'Sullivan, D. and T. J. Dunkerton (1995). Generation of inertia–gravity waves in a simulated life cycle of baroclinic instability. *J. Atmos. Sci. 52*(21), 3695–3716.

Pavelin, E., J. A. Whiteway, and G. Vaughan (2001). Observation of gravity wave generation and breaking in the lowermost stratosphere. *J. Geophys. Res.: Atmos. 106*(D6), 5173–5179.

Posselt, D. J. and C. H. Bishop (2018). Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter. *Quart. J. Roy. Meteorol. Soc. 144*(716), 2331–2349.

Satterfield, E., D. Hodyss, D. D. Kuhl, and C. H. Bishop (2017). Investigating the use of ensemble variance to predict observation error of representation. *Mon. Weath. Rev. 145*(2), 653–667.

Sawada, Y., K. Okamoto, M. Kunii, and T. Miyoshi (2019). Assimilating every-10-minute Himawari-8 infrared radiances to improve convective predictability. *J. Geophys. Res.: Atmos. 124*(5), 2546–2561.

Shlyaeva, A., J. S. Whitaker, and C. Snyder (2019). Modelspace localization in serial ensemble filters. *Journal of Advances in Modeling Earth Systems 11*, 1627–1636.

Simmons, A. J. and C. Temperton (1997). Stability of a two-time-level semi-implicit integration scheme for gravity wave motion. *Mon. Weath. Rev. 125*(4), 600–615.

Whitaker, J. S. and T. M. Hamill (2012). Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Weath. Rev. 140*(9), 3078–3089.

Zhang, F., M. Minamide, and E. E. Clothiaux (2016). Potential impacts of assimilating all-sky infrared satellite radiances from GOES-R on convection-permitting analysis and prediction of tropical cyclones. *Geophys. Res. Let. 43*(6), 2954–2963.

Zhang, Y., F. Zhang, and D. J. Stensrud (2018). Assimilating all-sky infrared radiances from GOES-16 ABI using an ensemble Kalman filter for convection-allowing severe thunderstorms prediction. *Mon. Weath. Rev. 146*(10), 3363–3381.

Zhu, Y., E. H. Liu, R. Mahajan, C. Thomas, D. Groff, P. van Delst, A. Collard, D. Kleist, R. Treadon, and J. Derber (2016). All-sky microwave radiance assimilation in the NCEP's GSI analysis system. *Mon. Wea. Rev. 144*, 4709 – 4735.