Use of cloud condensate in the background error formulation

Elías Valur Hólm[1] **and Jiandong Gong**[1,2]

(1) ECMWF, Shinfield Park, Reading
(2) National Meteorological Center, Beijing, China
Elias.Holm@ecmwf.int

1 Introduction

The main cloud observations used by weather forecasting centres are indirect measurements; they are in the form of top of atmosphere outgoing infrared and microwave radiances which are affected by a whole column of the atmosphere and the surface. The main difficulty in using radiance observations in a data assimilation system is that radiances are related to the model's state variables through a complex radiative transfer model. The radiative transfer model integrates a model column into a single number for comparison with the observed radiance - this process is called an observation operator. Conversely, when a radiance observation implies a change in the atmospheric state, a single number is distributed into updates to all those variables in the model column which affect the radiance. How accurately each model variable is updated depends on the accuracy of the observation operator, the background state and the estimated observation and background errors. In particular, if the background errors are not correctly estimated, then the signal can be attributed to the wrong variables. To give an example, specifying a humidity background error that is too large could cause radiance information on temperature and humidity to be excessively allocated to humidity. Accurate estimates of the background errors are thus essential to correctly attribute radiance observational information, in particular in cloudy and precipitating areas where the uncertainty is larger than in clear sky. At present no cloud variables are accounted for in the ECMWF analysis and this clearly increases the uncertainty in cloudy and precipitating areas. In this paper we will look at ECMWF's current background errors for water vapour and explore how it can be extended to include cloud variables.

2 Cloud background errors in context

Currently the linearized version of the radiance observation operators used in data assimilation at ECMWF takes prognostic temperature and humidity as input. It then diagnoses the cloud variables and precipitation fluxes needed in the calculation of model equivalents of the observed radiances. With this approach, temperature and humidity are updated by the assimilation system, but the initial condition of cloud variables is left unchanged. This approach has two significant limitations. First, errors in cloud variables may be wrongly interpreted as errors in humidity and temperature, because the linearized observation operators do not have prognostic cloud variables as input. Second, the forecast model may have to adjust the cloud variables to the changes in temperature and humidity through a spin-up process.

A more accurate approach to the assimilation of cloud sensitive observations is to also include prognostic cloud variables as input to the linearized observation operator and update the cloud variables in the initial conditions along with humidity and temperature. This requires developments in three areas:

1. Use of prognostic cloud variables in cloud sensitive observation operators, in particular cloudy RTTOV.

2. Developments to include cloud variables in the linear physics used by the data assimilation.
3. Development of background errors for cloud variables.

At ECMWF developments in all three areas are taking place in a concerted effort to make better use of cloud sensitive observations, in particular radiances. With this work we want to be able to answer two related questions:

- Does inclusion of prognostic cloud variables as input to the observation operator make a difference to the forecast impact of the data?
- Does updating the initial conditions of cloud variables make a difference to the forecast of clouds and precipitation?

3 Choice of variables for the cloud analysis

There are three different sets of variables in the analysis: nonlinear model variables, tangent linear model variables and control variables (see Table 1). The current (June 2010) nonlinear forecast model at ECMWF has three variables that together describe the evolution of water in the atmosphere: water vapour q_v, cloud condensate q_c and cloud fraction N. At the end of the physics in each timestep cloud condensate is split into cloud liquid water q_l and cloud ice q_i, which are the advected and stored model variables. The amount of liquid and ice is a function of temperature only,

$$q_l = \alpha(T)q_c \tag{1}$$

$$q_i = (1-\alpha(T))q_c \tag{2}$$

$$\alpha(T) = \begin{cases} 1 & \text{if} \quad T \geq T_{00} \\ ((T-T_i)/(T_{00}-T_i))^2 & \text{if} \quad T_{00} > T > T_i \\ 0 & \text{if} \quad T \leq T_i \end{cases} \tag{3}$$

where $T_{00} = 273.16K$ and $T_i = T_{00} - 23K$. In the current tangent linear model however, water vapour is the only variable describing the evolution of water. This difference is mainly due to the difficulty to accurately describe the dependency of cloud sensitive observations on cloud processes. This difficulty has made it more accurate to ignore changes to the initial conditions of all cloud and precipitation variables in the assimilation process and only update water vapour. With water vapour and temperature as the only prognostic thermodynamically active variables in the tangent linear model, diagnostic physics parameterization are used to calculate cloud and precipitation variables for use in linear physics and observation operators. As a result, cloud and precipitation sensitive observations project their information content onto temperature and water vapour. Water vapour is also the only water variable in the control variable. The control variables are linear combinations of the tangent linear variables, which are chosen so that the cross-correlation between the background errors in different variables is reduced as much as possible within the constraints of the covariance model used. The background error covariance matrices are then expressed in terms of the control variables, which are assumed uncorrelated.

Before adding new cloud tangent linear and control variables the main consideration is if these new variables help to extract observational information. The model and observation operators accuracy, sensitivity and linearity for these additional variables is a key point. We consider to start with the current version of the ECMWF cloud scheme, where water vapour, cloud condensate and cloud cover are the prognostic variables. Cloud condensate is a more fundamental variable than cloud cover, because cloud cover can be diagnosed quite accurately from the cloud condensate, e. g. by the Smith scheme (Smith, 1990) as has been verified by observations (Wood and Filed, 2000). There is also a fairly accurate way

Variables	Current	Planned
Nonlinear	T, q_v, q_c, N $q_l = \alpha(T)q_c$ $q_i = (1-\alpha(T))q_c$	$T, q_v, q_l, q_i, N, q_r, q_s$
Tangent linear	$\delta T, \delta q_v$	$\delta T, \delta q_v, \delta q_c$ $\delta q_l = \alpha(T^b)\delta q_c + \delta\alpha q_c^b$ $\delta q_i = (1-\alpha(T^b))\delta q_c - \delta\alpha q_c^b$
Control	$\delta T_u, (\frac{\delta q_v}{q_{sat}^b})_u$	$\delta T_u, (\frac{\delta q_v}{q_{sat}^b})_u, (\frac{\delta q_c}{f(T^b, q_v^b, q_c^b, N^b)})_u$

Table 1: Current (June 2010) and planned thermodynamic variables in the ECMWF IFS system. In the current nonlinear and planned tangent linear model cloud condensate is diagnostically split after the physics into cloud liquid and ice, the advected variables. The nonlinear extension of the cloud variables takes place in November 2010, but the tangent linear and control variable extensions are under development.

to split cloud condensate into cloud liquid and cloud ice as a function of temperature. For these reasons cloud condensate would be preferred over cloud liquid and ice or cloud cover. Another very practical reason to prefer cloud condensate over cloud cover in the analysis is that the processes governing cloud cover evolution are more nonlinear than those governing cloud condensate evolution. This makes it easier to include prognostic cloud condensate information accurately in the linear physics which then increases the accuracy of projecting cloud information from observations onto the model variables. Apart from these practical considerations, there are scientific reasons for starting with cloud condensate. Cloud condensate forms from humidity, and precipitation forms from cloud condensate, so one needs to include accurate cloud condensate before considering precipitation. For a further discussion of the different considerations that need to be made when choosing a cloud control variable see the review of Lopez (2006).

Adding cloud condensate to the analysis will make a distinct change to the treatment of water in the linear inner loops of the four-dimensional variational data assimilation (4D-VAR). In the current linear model, all water is lost from the system once it condenses, because there is no cloud condensate variable. Water vapour is (mostly) limited by condensation to $\frac{q_v}{q_{sat}} < 1$ (except when supersaturated with respect to ice). Similarly, cloud condensate is (mostly) limited by autoconversion (to precipitation) to $\frac{q_c}{Nq_c^{crit}} < 1+\varepsilon < 2$ (except in very strong convection). When cloud condensate is included in the linear system, the new frontier now becomes precipitation, where water is again lost whenever there is precipitation due to there not being any linear precipitation variable. Future developments will doubtless include precipitation in the analysis as well.

4 The water vapour control variable link to clouds

Before going on to the cloud condensate control variable, we will first look at how the current water vapour control variable relates to clouds. An important part of the current analysis is to account for the correlation between water vapour and temperature errors. The water vapour-temperature correlation $Q_{qT}(rh^b)$ is a function of relative humidity (Hólm et al. 2002) and is close to 1 at saturation and reduces to zero at about 80% relative humidity. It thus appears the correlation is mainly describing condensation and evaporation inside clouds. In fact, a plot of the correlation coefficient $Q_{qT}(rh^b)$ looks very similar to

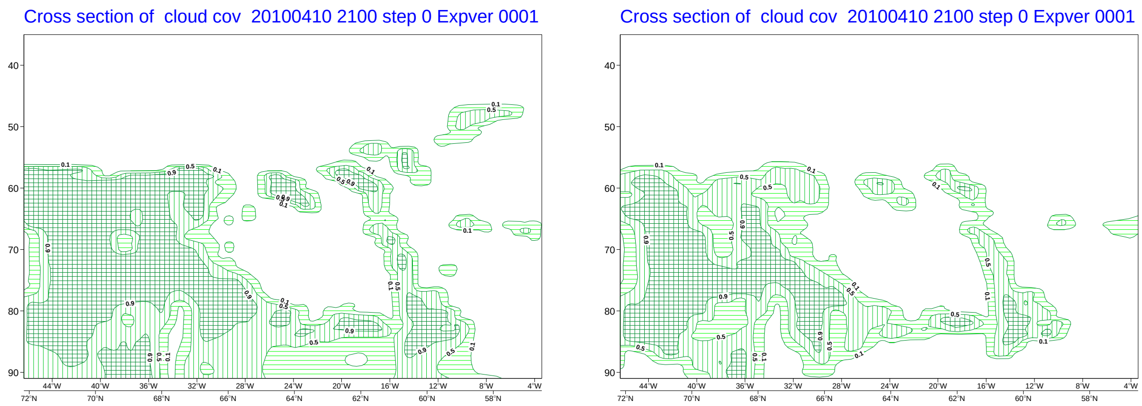

Figure 1: Cross section Greenland-Iceland of (left) Water vapour-temperature correlation $Q_{qT}(rh^b)$ in terms of $N_{eff} = \frac{q_v^b}{q_s(T^b)} Q_{qT}(rh^b)$. (right) Model first guess cloud cover N. The fields are similar, but the correlation also picks up additional processes, so $N_{eff} > N$.

a plot of cloud cover versus relative humidity for stratiform clouds, as e. g. in Tompkins and Janisková (2004). In a very simplified framework, neglecting entrainment/detrainment and precipitation effects, the in-cloud total water q_t^c is conserved,

$$\delta q_t^c = \delta q_v^c + \delta q_l^c + \delta q_i^c = 0 \tag{4}$$

where δq_l^c and δq_i^c are the in cloud liquid and ice increments. Considering a single gridbox with cloud cover N, the total water change can be estimated (see e. g. Bechtold and Cuijpers 1995) from the gridbox mean parameters and increments ($\delta q_v, T^b$ etc.) as

$$\delta q_t^c = \delta q_s(T^b) + \frac{\delta q_l + \delta q_i}{N} = \delta q_s(T^b) - \frac{\delta q_v}{N} = 0 \tag{5}$$

where we use $\delta q_v = \delta q_v^c$ since no change in water vapour takes place outside the cloud. This gives the gridpoint mean water vapour change in response to a temperature change as

$$\delta q_v = N \delta q_s(T^b) = N \left. \frac{\partial q_s}{\partial T} \right|_{T^b} \delta T \tag{6}$$

The water vapour control variable definition contains a similar relationship, after multiplying with $q_s(T^b)$ (Hólm et al. 2002),

$$\delta q_v = (\delta q_v)_u + Q_{qT}(rh^b) \frac{q_v^b}{q_s(T^b)} \left. \frac{\partial q_s}{\partial T} \right|_{T^b} \delta T \tag{7}$$

The 'balanced' part of the water vapour control variable would correspond to the simple description of cloud condensation effects if the correlation coefficient was related to cloud cover as

$$N_{eff} = \frac{q_v^b}{q_s(T^b)} Q_{qT}(rh^b) \approx N \tag{8}$$

Plotting the correlation coefficient shows there is indeed a close connection with the background cloud cover N, except close to the surface where $Q_{qT}(rh^b)$ is much larger than implied by the cloud cover. This indicates other effects than cloud condensation are contributing to the water vapour-temperature correlation close to the surface. The water vapour-temperature correlation is thus a sum of the cloud condensation part discussed above and other effects like convection and probably divergence close to the surface.

5 The cloud condensate control variable link to water vapour and temperature

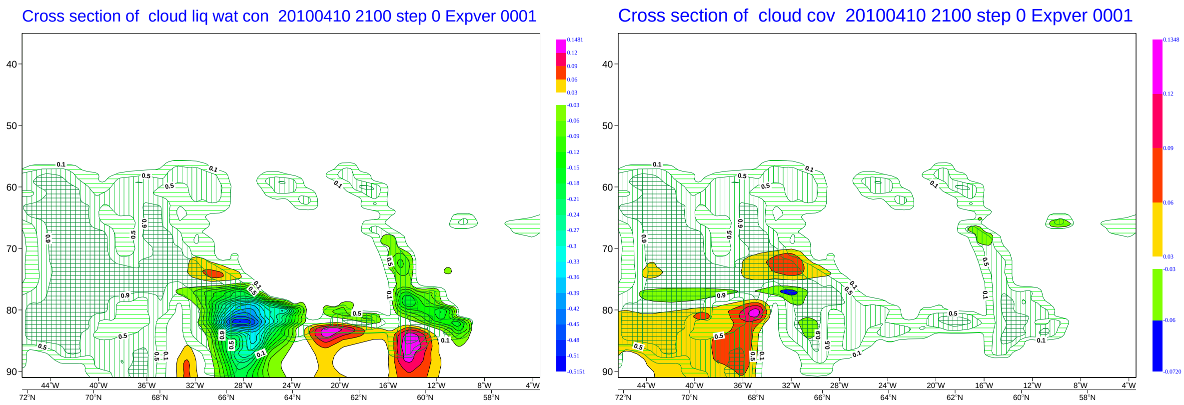

Figure 2: Cross section Greenland-Iceland (model first guess cloud cover in green) (left) Balanced δq_l increments implied by $Q_{qT}(rh^b)$. (right) Balanced δq_i increments implied by $Q_{qT}(rh^b)$. The implied increments are too large here because, because $Q_{qT}(rh^b)$ is just an approximation to the cloud condensate correlations which will be improved upon in subsequent research.

From the simplified cloud scheme considered above, one can see that changes in cloud condensate go along with changes in temperature and water vapour through total water conservation. This implies that cloud condensate changes can be expressed either in terms of humidity or temperature increments,

$$\delta q_c = -\delta q_v == -N \left.\frac{\partial q_s}{\partial T}\right|_{T^b} \delta T \tag{9}$$

As for humidity above, it is likely that the balance relationships for cloud condensate errors will follow a similar form and include both a water vapour and temperature term,

$$\delta q_c = (\delta q_c)_u - Q_{cT}(rh^b) \left.\frac{\partial q_s}{\partial T}\right|_{T^b} \delta T - Q_{cq}(rh^b)\delta q_v \tag{10}$$

Here $Q_{cT}(rh^b)$ and $Q_{cq}(rh^b)$ are regression coefficients. However, because temperature and water vapour are already correlated, this overlap needs to be accounted for by the regression. As an illustration of the cloud condensate balance, if all effects were accounted for by the temperature term, then the balance would be of the same form as between water vapour and temperature, just with a reversed sign. We apply this in Fig. 2, assuming $Q_{cT}(rh^b) = -Q_{qT}(rh^b)$. A straight application of this relationship gives large balanced cloud increments due to that N_{eff} is larger than N, especially close to the surface as seen in Fig. 2, where the negative liquid increment is larger than the total amount of liquid available in the first guess (not shown). However, the cloud condensate correlation with the other variables will pick up different effects in addition to condensation, and this means that one needs to derive $Q_{cT}(rh^b)$ and $Q_{cq}(rh^b)$ and then decide on the best form for the balance relationship.

When cloud condensate is used as control variable, one additional step between the balance operator and the tangent linear model is to convert cloud condensate increments back to cloud liquid and ice water increments, which are the variables advected by the tangent linear model. This conversion follows the same temperature dependent split as above for the nonlinear variables, possibly with a requlарization of α,

$$\delta q_l = \alpha(T^b)\delta q_c + \delta\alpha q_c^b \tag{11}$$

$$\delta q_i = (1-\alpha(T^b))\delta q_c - \delta\alpha q_c^b \tag{12}$$

6 Cloud condensate control variable candidates and their Gaussianity

A final criterion for choosing the control variable is that the probability density function (pdf) of the errors should be as close to Gaussian as possible. The reason for this is that the quadratic cost functions employed in data assimilation would be an exact model of the errors if they were Gaussian. Any deviation from Gaussianity in the background error pdf's reduces the accuracy of the background error model. For water vapour, Gaussianity was achieved by a state dependent normalization $f_q(rh)$ that depends on the mean of the background state and the background with the increments added, e. g. using $rh^b + \frac{1}{2}\delta rh$ instead of rh^b in the normalization (Hólm et al. 2002). For cloud condensate we can also experiment with different state dependent normalizations. The normalization can be a function of several of the variables affecting clouds, and research is needed to find the most appropriate form $f(T^b, q_v^b, q_c^b, N^b)$. In general, all regression parameters will be a function of model level L (or sigma coordinate) as well.

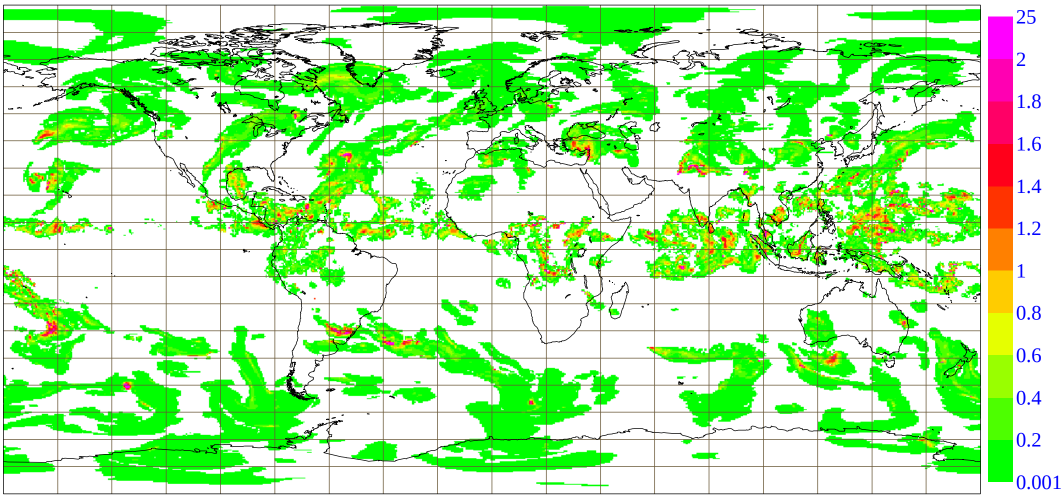

Figure 3: Normalized cloud condensate $\frac{q_c}{Nq_c^{crit}}$ level 60 ($\approx$ 400 hPa). The red areas are most likely precipitating.

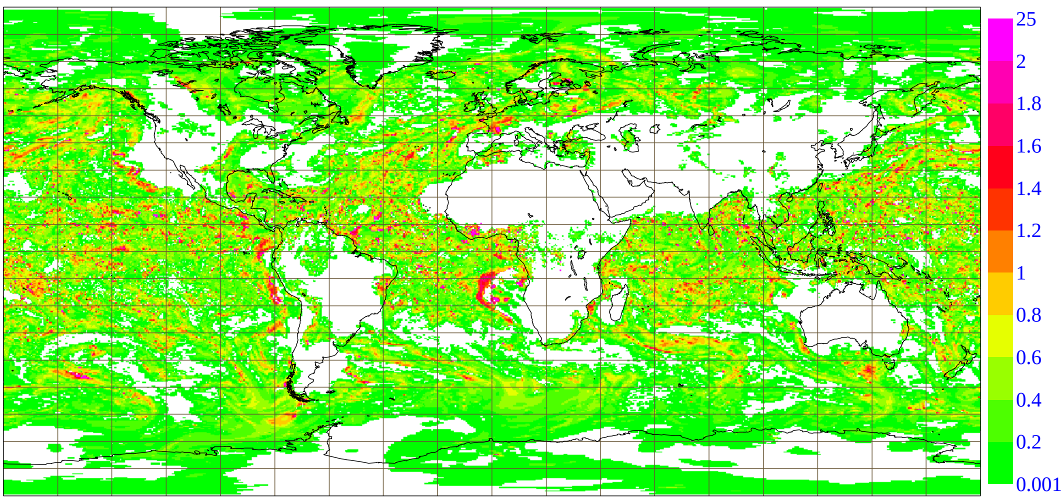

Figure 4: Normalized cloud condensate $\frac{q_c}{Nq_c^{crit}}$ level 80 ($\approx$ 900 hPa). The red areas are most likely precipitating.

As an initial investigation, we have studied samples of forecast differences between sets of 3-h forecasts from independent analyses using perturbed observations. The pdf's of the differences can be compared with a Gaussian reference distribution, and the closer the pdf is to Gaussian, the better the control variable candidate is likely to perform in the analysis. Figures 3– 4 show normalized cloud condensate $(q_c/N)/q_c^{crit}$ at $\approx$ 400 hPa and $\approx$ 900 hPa model levels. Looking at these fields it is obvious that there will be some problems in forming forecast difference statistics, as large areas have no cloud condensate. For this reason, the forecast difference sample is limited to points where $N > 0.01$. The normalized cloud condensate also shows some large values (red). Whenever $(q_c/N)/q_c^{crit} > 2$ there is most likely heavy convective precipitation, and these points are also excluded from the sample. We look at four candidates for the control variable:

1. $\delta q_c/\sigma(L)$
2. $\delta q_c/\sigma(L,rh)$
3. $\frac{\delta q_c}{N q_c^{crit}}/\sigma(L)$
4. $\frac{\delta q_c}{N q_c^{crit}}/\sigma(L, \frac{q_c}{N q_c^{crit}})$

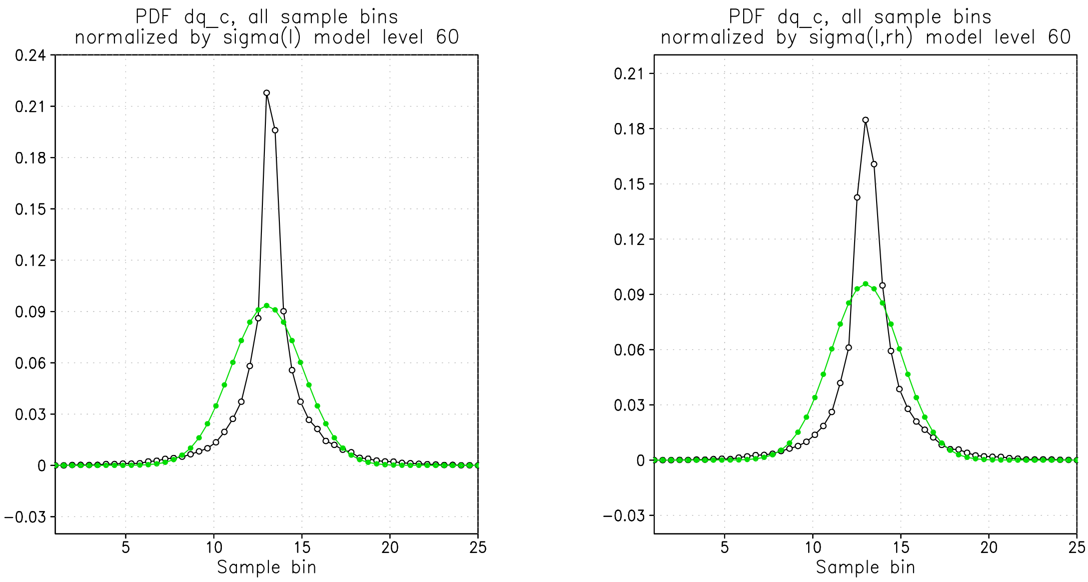

Figure 5: Cloud condensate δq_c: model level 60 ($\approx$ 400 hPa, mostly ice). Left: normalized by constant $\sigma(L)$ non-Gaussian, inhomogeneity causes relatively smaller values to accumulate close to zero. Right: normalized by flow dependent $\sigma(L,rh)$ still bad.

As can be seen in Figs. 5– 6 each successive change brings the pdf closer to a Gaussian at model level 60 ($\approx$ 400 hPa, mostly ice), but there is still some way to go even for the best candidate control variable. Figure 7 shows that model level 80 ($\approx$ 900 hPa, mostly water) has more Gaussian statistics. Further research will be performed along these lines to find a cloud condensate control variable with more Gaussian distribution.

7 Conclusion

In this paper we have presented normalized cloud condensate as a candidate cloud control variable to be used in the analysis. We also showed how the correlation of cloud condensate with water vapour and temperature could be included as a 'balanced' part of the control variable. The exact formulation of the normalization and the balance is still under active investigation. The next steps following on from the

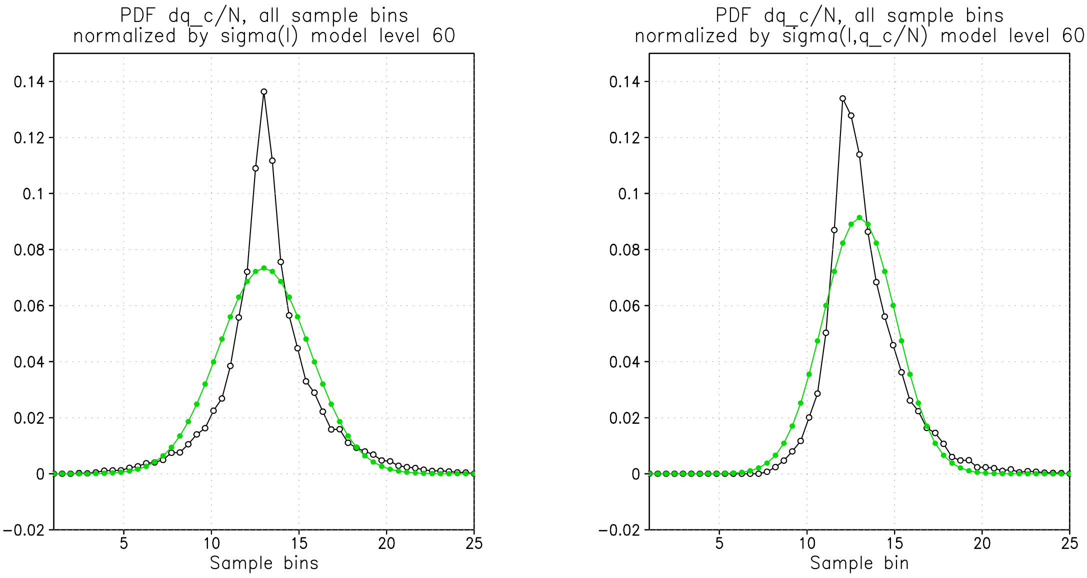

Figure 6: Normalized cloud condensate $\frac{\delta q_c}{N q_c^{crit}}$: level 60 ($\approx$ 400 hPa, mostly ice). Left: normalized by constant $\sigma(L)$ Right: normalized by flow dependent $\sigma(L, \frac{q_c}{N q_c^{crit}})$. Both similar and better than δq_c. Only include samples for $N > 0.01$ and $\frac{q_c}{N q_c^{crit}} < 2$.

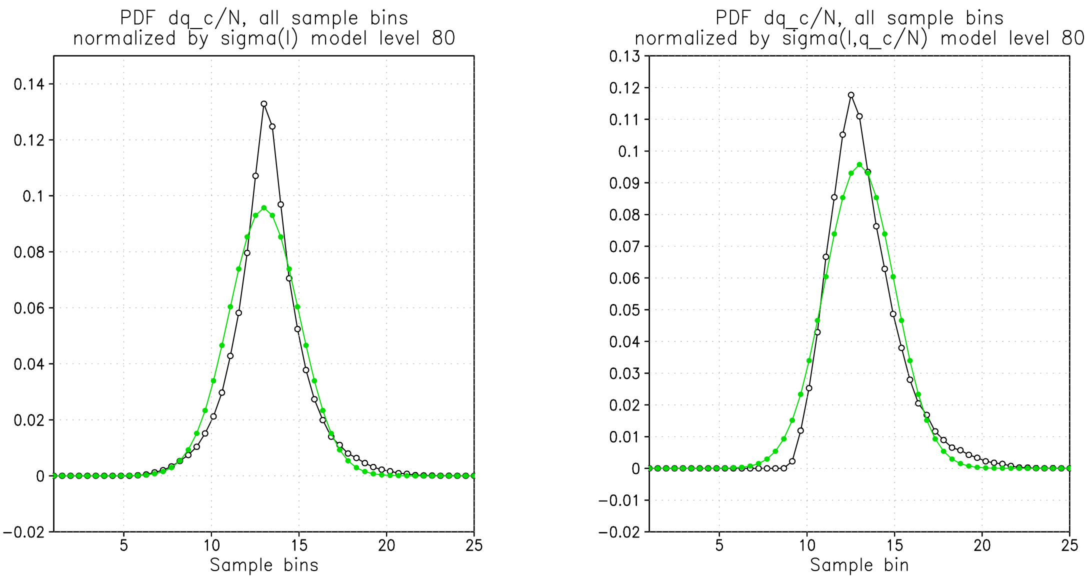

Figure 7: Normalized cloud condensate $\frac{\delta q_c}{N q_c^{crit}}$: level 80 ($\approx$ 900 hPa, mostly water). Left: normalized by constant $\sigma(L)$ Right: normalized by flow dependent $\sigma(L, \frac{q_c}{N q_c^{crit}})$. More Gaussian than upper (ice) levels.

current work will be to study how the background error formulation interacts with the updated linear physics including prognostic cloud condensate and how the whole system including cloud sensitive radiances performs.

References

[1] Bechtold, P., and J. W. M. Cuijpers, 1995: Cloud perturbations of temperature and humidity: a LES study. *Boundary-Layer Meteorology*, **76**, 377-386.

[2] Hólm, E., E. Andersson, A. Beljaars, P. Lopez, J.-F. Mahfouf, A. J. Simmons, and J.-N. Thépaut, 2002: Assimilation and Modelling of the Hydrological Cycle: ECMWFs Status and Plans. *ECMWF Technical Memorandum 383.*

[3] Lopez, P., 2006: Cloud and precipitation parameterizations in modeling and variational data assimilation: a review. *J. Atmos. Sci.*, **64**, 3766-3784.

[4] Smith, R. N. B, 1990: A scheme for predicting layer clouds and their water content in a general circulation model. *Q. J. R. Meteorol. Soc.*, **116**, 435-460.

[5] Tompkins, A. M., and M. Janisková, 2004: A cloud scheme for data assimilation: Description and initial tests. *Q. J. R. Meteorol. Soc.*, **130**, 2495-2517.

[6] Wood, R., and P. R. Field, 2000: Relationship between total water, condensed water, and cloud fraction in stratiform clouds examined using aircraft data. *J. Atmos. Sci.*, **57**, 1888-1905.