Challenges in moisture assimilation for the convective-scale

March 24, 2016

1 Introduction

The importance of data assimilation (DA) has been appreciated for as long as numerical weather prediction (NWP) models demanded accurate initial conditions [Daley (1991)]. One of the first problems that DA had to deal with, and which is an historical parallel to the problems faced today with moisture, was in the generation of spurious gravity waves by the DA itself which swamped models with noise. One strategy to deal with this problem was to damp the gravity waves from the affected analyses. This process, called initialization (see Ch. 6 of [Daley (1991)]), did reduce these shocks, but produced initial conditions that were almost inevitably less in agreement with the observations than was the uninitialized analysis. A better solution to this problem was found to be on-line filtering of gravity waves, which is achieved in DA via either an appropriate forecast error probability density function (PDF), described presently with a background error covariance matrix (or with an on-line initialization term - known as the ’Jc’ term [Courtier et al. (1998)]). The relative sizes of the balanced and unbalanced (gravity) corrections that are permitted to the background state are specified by the background error covariance matrix, which helps to ensure that an analysis is produced without excessive gravity waves and at the same time giving a good fit to the observations.

Analogous problems exist for the moisture analysis in the latest family of NWP models, which are capable of permitting convection explicitly. The Met Office’s current operational convective-scale model for instance has variable resolution over the United Kingdom domain and has a core grid of resolution 1.5km (the model is known as the ’UKV’ - the United Kingdom Variable resolution model [Tang et al. (2013)]). In these high-resolution models, there is scope for errors in the initial conditions to lead to spurious forecast features that are associated with moist processes. Such features include excessive convection or precipitation, especially at the early stages of a model run (see e.g. [Holm et al. (2002)] for the case of the ECMWF system in the tropics, and [Dixon et al. (2009)] in the case of a 4km gridlength version of the UM over the UK). This property of forecasts is often called the ’spin-down’ problem. Further consequences of this problem are discussed in Sect. 4.

The main cause of the gravity wave problem is thought to be due to lack of accuracy of data [Ghil (1980)], but in convective-scale systems there is the additional problem of systematic model error, due to our inability to treat moist processes in models correctly. We may therefore assume that even if highly accurate initial conditions are provided from the DA, a model like the UKV will still suffer with initial shocks. In this case this is not because the initial conditions contain errors respect to the truth, but because they contain inconsistencies with respect to the model. The solution to the DA problem may therefore involve improvements to the background error covariance matrix that not only (i) represent realistic moist processes, but also (ii) discourage model increments that may lead to moisture-related instabilities in the model, even though these increments may have drawn the analysis closer to the truth.

Is it not the role of Four-dimensional variational data assimilation (4D-Var) to produce analyses with consistent and realistic trajectories? The answer is ‘yes possibly’, but there are some reasons why we should not rely on 4D-Var to solve this problem. Firstly, moist processes are typically absent (wholly or partially) in the linear models of 4D-Var [Rawlins et al. (2007)]. Secondly, explicit moist
Convection has significant non-linear behaviour over the typical analysis time window of a convective-scale system of a few hours, so in linear models where moist processes are present, the linearity requirements of 4D-Var can be violated, even in synoptic-scale models [Stiller and Ballard (2009)]. This can project observational information (especially from the end of the time window) backwards incorrectly to effectively cause the aliasing of model error on to the initial conditions. Thirdly, 4D-Var should not ideally be used to compensate for problems in the background error covariance matrix at the start of the time window. The emphasis on this paper is on the background error covariance matrix at the beginning of the time window, which is hoped will lead to improvements whether it is used in 3D or 4D-Var.

This paper reviews recent advances in moisture assimilation, which highlights the challenges that moisture assimilation systems face and the large number of routes that can be followed when designing new systems for high-resolution models. The paper then goes on to suggest a possible path that may be taken by the United Kingdom’s Met Office and other centres.

2 Current and upcoming observations of moisture and precipitation

Observations of moisture are extremely important for NWP. [Mahfouf (2011)] identifies three different kinds of observing system used in NWP, namely (a) conventional observations (measurements made in-situ), (b) ground-based remotely-sensed observations (measurements made from radar and profil-ers), and (c) space-based remotely-sensed observations (measurements made from satellite imagers and sounders). All three potentially contain valuable information about moisture in the atmosphere, but the remotely-sensed observations require a degree of interpretation by the assimilation system, which can cause problems if this is not done correctly (see Sect. 4). This is an issue especially in precipitating and cloudy regions, where important diabatic processes are happening, making forecasts particularly sensitive to initial conditions [Ohring and Bauer (2011)], but where observations from space are difficult to use [McNally (2002), Bauer et al. (2011)]. Even observations of non-humidity-related quantities like temperature and wind have the potential to influence the moisture field (via cycling in 3D-Var and additionally via the model within 4D-Var), but we focus here on observations that have a more direct relationship to the moisture fields of the model.

2.1 In-situ observations

Most in-situ observations of moisture measure water vapour and are in the form of relative humidity or dew-point temperature. Table 1 gives a selection of sources of ground-based observations, which include near-surface-level instruments (such as from one of the 200 or so surface stations operated by the Met Office, and from ships and buoys), and from radio-sondes on weather balloons. Ground-based observation are still important in NWP [Ingleby (2015)]. Humidity data are sometimes made from commercial aircraft via the AMDAR system, which the Met Office plans to assimilate in the future [Ingleby et al. (2013)]. Of all the data sources, in-situ measurements are thought to be the most accurate as their interpretation is relatively simple (minimising scope to produce artifacts in the analysis) and require the least pre-processing (although they are not without issues, see e.g. [Miloshevich et al. (2001)]).

2.2 Ground-based remotely-sensed observations

Convective-scale models are increasingly relying on remotely-sensed data collected from ground stations. Table 2 lists the main sources of ground-based remotely-sensed data. Radar reflectivity - yielding precipitation data - are especially important at the moment as water and ice hydrometeors can be inferred from these measurements. It remains a challenging task to assimilate precipitation
data within a variational framework owing to the need for highly non-linear micro-physical processes to be represented by 4D-Var and its potential expansion to allow hydrometeors to be available to the observation operator. Workarounds like latent heat nudging (Sect. 3.3) are currently used instead although centres are working towards the variational assimilation route [Sun et al. (2014)].

### 2.3 Space-based remotely-sensed observations

A wide variety of observations sensitive to water are made from space, but only a subset of instruments are currently employed with the UKV model. Table 3 lists a section of sources of space-based observations, which comprise infrared (IR) and microwave (MW) sounders and GPS radio occultation instruments. IR and MW data each bring their own benefits to NWP [Bauer et al. (2011)], e.g. IR data provide information about water vapour in various vertical layers of the atmosphere, but are more difficult to interpret over cloudy scenes, and MW data are easier to use in cloudy scenes, but give information on the total column amounts of water, and require more complicated radiative transfer model which account for multiple scattering effects. There are also issues with the assimilation of data that can 'see' the surface, where knowledge of surface emissivity of the ocean is more accurate than over land.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Type</th>
<th>Horiz. res. (km)</th>
<th>Sensitive to</th>
<th>Refs</th>
<th>UKV assim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface stations</td>
<td>point</td>
<td></td>
<td>RH, dew-point temperature</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Radio-sondes</td>
<td>point</td>
<td></td>
<td>RH, dew-point temperature</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>AMDAR, TAMDAR</td>
<td></td>
<td>RH</td>
<td></td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1:** Sources of in-situ observations of moisture. Note the following acronyms: AMDAR (Aircraft Meteorological Data Relay), TAMDAR (Tropospheric Airborne Meteorological Data Reporting).

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Type</th>
<th>Horiz. res. (km)</th>
<th>Sensitive to</th>
<th>Refs</th>
<th>UKV assim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar</td>
<td>Reflectivity</td>
<td></td>
<td>Precipitation</td>
<td>LHN</td>
<td></td>
</tr>
<tr>
<td>Radar</td>
<td>Refractivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GNSS</td>
<td></td>
<td>40</td>
<td>Total column water vapour</td>
<td>[Bender et al. (2011), Bennitt and Jupp (2012)]</td>
<td>Yes</td>
</tr>
<tr>
<td>Celiometer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2:** Ground-based remote-sensing instruments with sensitivity to atmospheric moisture. Note the following acronyms: GNSS (Global Navigation Satellite Systems), LHN (Latent Heat Nudging).
<table>
<thead>
<tr>
<th>Instrument</th>
<th>Type</th>
<th>Platform</th>
<th>WV channels</th>
<th>Horiz. res. (km)</th>
<th>Sensitive to</th>
<th>Ref(s)</th>
<th>UKV assim</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMSU-A</td>
<td>MW sounder</td>
<td>METOP-A/B</td>
<td>23.8, 31.4, 50.3, 89.0</td>
<td>45</td>
<td>Total column WV</td>
<td>[Derber and Collard (2011)]</td>
<td></td>
</tr>
<tr>
<td>MHS</td>
<td>MW sounder</td>
<td>NOAA-18/19, METOP-A/B</td>
<td>89, 157 GHz</td>
<td>16</td>
<td>WV near surface</td>
<td>[Guan et al. (2011)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>183±1 GHz</td>
<td></td>
<td>WV upper tropo.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>183±7 GHz</td>
<td></td>
<td>WV mid tropo.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>190.3 GHz</td>
<td></td>
<td>WV lower tropo.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSM/I</td>
<td>MW sounder</td>
<td>DMSP</td>
<td>22.24, 37.0 GHz</td>
<td>10s</td>
<td>Total column WV</td>
<td>[Okamoto and Derber (2005)]</td>
<td>No</td>
</tr>
<tr>
<td>SEVIRI</td>
<td>IR sounder</td>
<td>Meteosat-8/9/10/11</td>
<td>6.2, 7.3 μm</td>
<td>4-6 (Europe)</td>
<td>WV mid tropo.</td>
<td>[Montmerle et al. (2007), Jin et al. (2008)]</td>
<td></td>
</tr>
<tr>
<td>HIRS</td>
<td>IR sounder</td>
<td>METOP-A/B</td>
<td>6.5, 7.3 μm</td>
<td>10</td>
<td>WV mid tropo.</td>
<td>[Montmerle et al. (2007), Derber and Collard (2011)]</td>
<td>No</td>
</tr>
<tr>
<td>AIRS</td>
<td>IR sounder</td>
<td>Aqua</td>
<td>1334.6, 1427.2 cm⁻¹, 1563.7 cm⁻¹</td>
<td>13.5</td>
<td>WV mid tropo.</td>
<td>[Parkinson (2003)]</td>
<td>No</td>
</tr>
<tr>
<td>IASI</td>
<td>IR sounder</td>
<td>METOP-A/B</td>
<td>1320.00, 1349.50, 1372.25, 1392.50, 1393.00, 1395.25, 1396.75, 1398.25, 1401.50</td>
<td>12</td>
<td>WV upper tropo.</td>
<td>[Collard (2007), Guidard et al. (2011)]</td>
<td>No</td>
</tr>
<tr>
<td>GPS-RO Cloud-Sat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Space-based remote-sensing instruments with sensitivity to atmospheric moisture and a selection of the relevant associated channels. Note the following acronyms: AMSU-A (The Advanced Microwave Sounding Unit-A), MHS (Microwave Humidity Sounder), SSM/I (Special Sensor Microwave/Imager), SEVIRI (Spinning Enhanced Visible and Infrared Imager), HIRS (High Resolution InfraRed Sounder), AIRS (Atmospheric InfraRed Sounder), IASI (Infrared Atmospheric Sounding Interferometer), MW (MicroWave), IR (InfraRed).
3 Methods of assimilation of moisture information

3.1 Moisture data assimilation for models like the UKV

Generally speaking there are many decisions that have to be made in order to facilitate new kinds of moisture observations for a particular system and these are likely to depend upon how the system is currently configured. With respect to the adaptation of an existing Var system, some of the issues are listed here.

- **What observations to assimilate?** This depends upon their availability, their ability to be modelled in the assimilation, their impact, and on their error characteristics. The second point requires careful consideration as the observation operator may need inputs that are unavailable in a particular system. An example is the assimilation of cloud or precipitation affected satellite radiances. Failing to account for the cloud and precipitation contributions to the radiance correctly in the observation operator will cause this error to project onto other variables like humidity and temperature to compensate. This and other examples are discussed further in Sect. 6.

- **3D-Var or 4D-Var?** 3D-Var is obviously the cheaper option, but 4D-Var will allow variables to be evolved within the assimilation window. The latter issue is very important when different variables (like cloud, temperature and humidity) are strongly coupled as 4D-Var will allow observations to update all relevant variables necessary for a consistent model trajectory, even if they are not directly observed. 4D-Var opens well-known problems of its own though due to the non-linear nature of moist processes [Xu (1996), Stiller and Ballard (2009)], including convergence problems.

- **How long an assimilation time window?** In synoptic-scale models the memory of moisture perturbations in the model is shorter than for other variables [Bengston and Hodges (2005)], and is expected to be very short in convection-permitting models. A short time window (e.g. sub one-hour, even as short as a few minutes) may allow fast processes to be approximated in a linear manner in 4D-Var (and may alternatively allow 3D-Var to be used as a reasonable approximation). It would though require the state error covariance information to be propagated from one assimilation cycle to the next, and may anyway be impractical due to the availability of observations for assimilation at such short notice. A longer time window (e.g. six hours or more) with 4D-Var will exacerbate the non-linearity problem mentioned above.

- **What analysis variables to use?** Conventionally data assimilation systems have a single moisture variable, but there is growing need for data assimilation systems to analyse variables more in line with the model’s requirements and with the information that is available from upcoming observations. This may mean introducing new variables related to cloud and precipitation. Cloud and precipitation may be diagnosed from the conventional variables, and their perturbations may be propagated within 4D-Var, but it is thought that using separate variables would be advantageous (e.g. [Gong and Holm (2011)]). The nature of each moisture analysis variable also requires thought in order to account for the multi-variate nature of background error covariances and for the inevitable non-Gaussianity of such variables. The solution to these problems is sought via the control variable transform method, Sect. 3.2, and is discussed further in Sects. 5 and 7.

- **Reduce imbalance via initialization?** Ideally the degree of imbalance introduced by the assimilation would be fully controlled by the background error covariances (here we refer to imbalance in a general sense as not just ageostrophic and non-hydrostatic balance, and so we include moisture imbalances - e.g. the addition of water, or the lowering of the temperature to cause excessive precipitation (spin-down)). The Met Office mitigates the effect of assimilation on imbalance by introducing analysis increments bit-by-bit over a time window (). This is done via a method called Incremental Analysis Update (IAU).
• Accompany the Var system with ad-hoc data assimilation of data? Existing observation systems sensitive to moisture are currently exploited even without many of the above modifications in place to use these data within Var. Two important ones in the Met Office are latent heat nudging and cloud nudging. These exploit pre-analyses of precipitation and cloud, and modify the analysis post assimilation as the IAU is run. These methods are discussed in Sect. 3.3.

These and further arguments are discussed with regard to ensemble data assimilation in Flowerdew (2013, unpublished) and examples of where moisture assimilation can 'go wrong' if data assimilation is not done appropriately are discussed in Sect. 4.

The UKV model itself has the ability to recognise six moist mixing ratio variables: water vapour, cloud water, rain water, ice, snow and graupel [Lean et al. (2008)]. In practice though only four are used as prognostic variables at present: water vapour, cloud water, rain water and a single combined ice and snow variable [Lean et al. (2008), Leoncini et al. (2013)]. Ideally data assimilation should allow all variables (including all of the active moisture variables) to be updated. At the time of writing the Met Office uses 3D-Var for the UKV assimilation, with three-hour cycling which produces analysis increments that comprise horizontal wind components ($\delta u$ and $\delta v$), potential temperature ($\delta \theta$), density (in the form $r_E^2 \delta \rho$ where $r_E$ is the Earth's radius), exner pressure ($\delta \Pi$), aerosol, and total water ($\delta q_T = \delta q_v + \delta q_l + \delta q_i$, where the components in the sum are mixing ratios of vapour, liquid and ice respectively). The total water increment is broken down into separate contributions by the UKV model itself. There is scope to modify this system by considering the various issues given above.

### 3.2 Background error covariance modelling in Var (control variable transforms)

For the last few decades Var has been the leading method of estimating initial conditions for NWP models, e.g. [Parrish and Derber (1992), Rabier et al. (1998), Derber and Bouttier (1999), Gauthier et al. (1999), Lorenc et al. (2000), Rawlins et al. (2007)]. Even though ensemble techniques have gained in popularity over recent years, Var (or at least Var hybridised with an ensemble system as in e.g. Clayton et al. (2012)) is still widely used, and will continue to be used for the foreseeable future. The success of Var for large-scale weather forecasting is thought to be due partly to the use of a background error covariance matrix ($B$) that includes physical processes (like dynamical balance) and realistic correlation length scales. Due to its large size the $B$-matrix is not stored explicitly, but is instead modelled by control-variable-transform-based (CVT) or diffusion-based methods (see [?] for a review). The CVT method is a powerful way of accounting for multivariate and spatial aspects of background errors, and appropriate adaption of the current scheme may well be the solution to many of the issues listed in Sect. 3.1 for the convective-scale. The CVT method revolves around a transform (denoted $U$) between pseudo model space increments ($\delta x_{pm}$)$^1$ and control space increments ($\delta \chi$). In this paper, this transform takes the following generic form:

$$
\delta x_{pm} = U \delta \chi,
$$

$$
= S^{-1} U_p \delta \tilde{\chi},
$$

$$
= S^{-1} U_p \Sigma U_s \delta \chi,
$$

(1)

where $\delta \tilde{\chi} = \Sigma U_s \delta \chi$ is an intermediate vector consisting of parameters (variables with a tilde in this paper denote parameters) and other symbols are defined below. The vectors $\delta x_{pm}$, $\delta \tilde{\chi}$ and $\delta \chi$ are different representations of the same increment. Background errors in the pseudo model space have covariance $\left\langle \delta x_{pm}^b, \delta x_{pm}^b \right\rangle = B_{pm}$, but the background errors in control space are assumed to have

---

$^1$For convenience we use the term "pseudo model space", rather than just "model space" to refer to a space that includes variables that may not be exactly model variables, but are the direct output of a CVT under examination. For example a particular CVT may output the logarithm of specific humidity (making this a pseudo model space variable), but the corresponding model space variable may be (non-logarithmic) specific humidity. Many variables though will be the same between the pseudo and actual model spaces.
covariance $\langle \delta \chi^b \delta \chi^{bT} \rangle = I$, where the “b” superscript indicates that the increments are possible samples of background error. Thus, background errors in control space are mutually uncorrelated and have unit variance. In (1) $U$ is the spatial part of the CVT (which introduces the spatial covariances within the set of control parameters), $\Sigma$ is the the matrix of standard deviations of parameters (if used in a particular DA system), $U_p$ is the parameter part of the CVT (which introduces multivariate relationships by transforming control parameters to pseudo model variables - this is sometimes called the 'balance operator'), and $S^{-1}$ is the inverse of a simplification operator (if used in a particular DA system; $S$ may e.g. reduce the resolution from the full model resolution).

Minimization is performed in control space. This involves writing an incremental cost function in terms of $\delta \chi$, $J_x[\delta \chi]$ (see e.g. [Bannister (2008b)]) and adjusting $\delta \chi$ to minimize $J_x[\delta \chi]$. This is mathematically equivalent to minimizing the incremental cost function in model space, $J_z[\delta x_{pm}]$, but with background error covariance $UU^T$, which is known as the implied background error covariance matrix. $J_x[\delta \chi]$ and $J_z[\delta x_{pm}]$ are related via $J_x[\delta \chi] = J_z[U \delta \chi]$.

For minimization purposes, $J_x[\delta \chi]$ is significantly simplified and numerically better conditioned than $J_z[\delta \chi]$. The special value of $\delta \chi$ that minimises $J_x[\delta \chi]$ is $\delta \chi^a$, which leads to the pseudo model space increment, $\delta x^a$, via (1), which is added to a reference state (e.g. the background) to give the analysis: $x^a = x^b + \delta x^a$.

The CVT method has been successful for large-scale systems where the physics that is being modelled is well-known (e.g. that hydrostatic balance is a very good approximation). Convective-scale models though pose new difficulties as the relevant balances (if any) are unclear [Berre (2000), Bannister et al. (2011)] and the forecast models have a greater need for information about moisture variables [Dance (2004), Fischer et al. (2005), Michel et al. (2011)].

The UKV assimilation system uses the following control parameters [Lorenc et al. (2000), Ingleby (2001), Bannister (2008b)]: $\delta \psi$ (streamfunction, for the rotational wind), $\delta \chi$ (velocity potential, for the irrotational wind), $\delta \rho$ (ageostrophic pressure), and $\delta \mu_{NLA}$ (a non-linear adaptive humidity variable). The additional geostrophic pressure, $\delta \rho$, is diagnosed from $\delta \psi$ via a mass-wind balance operator, $L_{hh}$, i.e. $\delta \rho = L_{hh} \delta \psi + \delta \rho_h$, the temperature, $\delta T$, is derived from the total pressure via the hydrostatic operator, $L_{hh}$, i.e. $\delta T = L_{hh} \delta \rho$, and total water, $\delta q_T$ is derived from $\delta \mu_{NLA}$ and $\delta T$ via the equation given later as (12). The particular choice of $U_p$ used follows from $L_{hh}$, $L_{hh}$, and (12). Section 3.2 of [Bannister (2008b)] shows how this is done and [Bannister (in preparation)] discusses the appropriateness of the balance relationships used for convective-scale flow.

This system of control variables and transforms will have to be modified in order to include, for instance, multivariate coupling between background errors of moisture and divergent wind, and to explicitly account for different phases of water in the assimilation.

### 3.3 Ad-hoc methods of assimilation of moisture information

Even with techniques like 4D-Var available for dynamical variables, some of the information about moisture is still often treated with ad-hoc techniques. One such method is latent heat nudging [Jones and Macpherson (1997)] (LHN). This is a method of assimilating surface precipitation observations, derived primarily from radar (e.g. as documented for the system used for the 2012 Olympics [Golding et al. (2014)]), which is still used at the Met Office given the absence of advanced precipitation assimilation within 4D-Var. The idea relies on the assumption that the latent heat release, integrated over a column of the atmosphere (levels 0 to $L$), is proportional to the precipitation rate at the bottom of that column. Applying this to a model integration of a background state (superscript b) at a particular geographical location:

$$\sum_{k=0}^{L} \delta \theta^b_k = \beta PR^b,$$

where $\delta \theta^b_k$ is the potential temperature increment due to moist parametrisations at level $k$ over the background trajectory, $PR^b$ is the precipitation rate forecast from that trajectory, and $\beta$ is a constant.
The aim of LHN is to derive a correction to $\delta \theta^b_k$ (called $\delta \theta^{LHN}_k$) which gives rise to the observed precipitation rate ($PR^b_{ob}$). This correction is assumed to have the form $\delta \theta^{LHN}_k = \alpha \delta \theta^b_k$, where $\alpha$ is constant over the column. Thus:

$$\sum_{k=0}^{L} (1 + \alpha) \delta \theta^b_k = \beta PR^b_{ob}. \quad (3)$$

Combining (2) and (3) leads to the following increment which is nudged into the model integration (in addition to variational analysis increments) using the IAU:

$$\delta \theta^{LHN}_k = \alpha \delta \theta^b_k = \frac{PR^b_{ob} - PR^b_{ob}}{PR^b_{ob}} \delta \theta^b_k, \quad (4)$$

(subject to certain conditions, e.g. $PR^b_{ob} > 0$).

Despite its usefulness (e.g. [Dixon et al. (2009)]), there are clear flaws to this procedure (e.g. the assumptions inherent in (2) and (3), neglect of advection of precipitation between neighbouring grid boxes, and the non-linear response from the model to $\delta \theta^{LHN}_k$), which motivate the development of a 4D-Var solution. Until recently, a procedure called MOPS (Moisture Observation Processing System, [Macpherson et al. (1996)]) was used in the Met Office to also nudge observations of cloud from satellite images and surface observations into the model. This is now done in 3D/4D-Var via a cloud fraction product ([Renshaw and Francis (2011)]). Other examples of diabatic initialization are discussed in [Sun et al. (2014)].

### 4 Implications of assimilating moisture (impact on other variables and things that can go wrong)

There are a number of problems, obstacles and research questions in the way of assimilating the full range of moisture-related observations in variational schemes in ways that are beneficial to the emanating forecast. There are also a number of known moisture-related problems that existed in the past, which are still noteworthy.

#### 4.1 'Leaky' structure functions

The smoothing effect of $B$ is known to be generally beneficial in data assimilation, but structure functions in $B$ that are too broad could result in un-physical analyses. This is especially true in regions of strong gradients such as the top of a boundary layer (BL) inversion or at the tropopause, and when the gradient is in the wrong place in the background forecast [Renshaw and Francis (2011), Ingleby et al. (2013)].

A pathological case is shown in [Ingleby et al. (2013)] for the case of a RH observation near the top of the BL and a similar picture to theirs is given as Fig. 1a. A saturated boundary layer (true RH field, $\mu$, black line) is forecast by the model to be deeper ($\mu^B$, blue line). An RH observation made just above the true BL, but just below the forecast one suggests that the model’s RH field should be reduced in the region defined by the RH structure function associated with this position (shape of the yellow curve). The fact that $\mu$ and $\mu^B$ have the sharp feature at the top of the boundary layer means that the local difference between the observation and the background is large. Adding the analysis increment ($\delta \mu$, yellow curve) to the background gives the analysis ($\delta \mu^A$, red curve). This results in a smoothed BL top and negative RH values above this, even though neither the background nor the observation values are themselves negative.

Another example is shown for temperature as Fig. 1b, which shows a BL top temperature maximum (black curve), again with the background forecasting a BL that is too deep (blue curve). Assimilating separately a temperature observation gives increments (yellow curve) which again smooths out the
Figure 1: Schema showing the damaging effect of assimilating data close to a sharp feature - in this case (a) RH and (b) temperature fields around the top of a saturated boundary layer inversion - when the background forecasts the feature in the wrong position. Here the boundary layer inversion top is forecast too high (blue) compared to the truth (black). The assimilation produces negative increments (yellow) to give the analysis (red). The observation is a point measurement of RH in (a) and temperature in (b) at each height where the increment curve peaks in magnitude.

features. If the background error covariances do not couple temperature and humidity correctly then this analysis may result in moisture imbalance (in this case a sudden drop in temperature leading to condensation), which the subsequent forecast will respond by precipitating the excess.

Problems like this are exasperated by inappropriate structure functions which can ‘leak’ information between regions of the atmosphere which would otherwise be expected to be largely decoupled. [Fowler et al. (2010)] discusses this issue further for the BL temperature problem with different kinds of BL found in ensembles of forecasts from the Met Office’s older NAE (North Atlantic and Europe) model. Importantly the problem with temperature is that the static stability can change anomalously as a result of the assimilation (in the case of Fig. 1b it has reduced, in a vain attempt to lower the BL top) which can then go on to affect cloud and precipitation forecasts. In neither of panels (a) or (b) is the BL depth reduced to its true depth so feature shifting algorithms are likely to be useful here, which can avoid large assimilation increments by displacing features instead, [Fowler et al. (2012)].

4.2 Assimilation of total column quantities in incompatible scenes

Total column water vapour (TCWV) measurements are naturally extracted from instruments like AMSU-A and SSM/I (Sect. 2). For a 1D column of the atmosphere the observation operator for TCWV is linear in specific humidity, $q$, and has the following form: $H = (\rho_1 \cdots \rho_L)$ when $x$ has the form $x = (q_1 \cdots q_L)$, where $\rho$ is the air density and the subscripts represent the vertical level indices. Assimilating a TCWV measurement ($y_{TCWV}^O$) updates the values of $q$ in each level in a way according to the following formula:

$$x^A = x^B + (HB)^T (R + HBH^T)^{-1} (y - Hx^B),$$

$$q^A_l = q^B_l + \sum_{l'=1}^{L} \rho_{ll'} B_{ll'} \left[ \frac{y_{TCWV}^O - y_{TCWV}^m}{\sigma_{TCWV}^O} + \sum_{l'=1}^{L} \rho_{ll'} B_{ll'lt'} \right],$$

where $y_{TCWV}^m = Hx^B$, $B_{ll'}$ is the error covariance between $q^B_l$ and $q^B_{l'}$, and $\sigma_{TCWV}^O$ is the expected observation error. Note that $B_{ll'} = \sigma_{ll'}^B \sigma_{ll'}^B$, where $\sigma_{ll'}^B$ is the expected background error in $q^B_l$ and
Figure 2: Frequency distributions of RH in the UKV for five days of three-hour forecast data from 1st to 5th September 2013 (validity time 15Z). Panel (a) is aggregated over the horizontal domain and for levels 1-16 (boundary layer), panel (b) is for levels 17-52 (free troposphere), and panel (c) is for levels 53-65 (upper troposphere and lower stratosphere). The topmost five levels of the model data are discarded.

\[ r_{\ell'\ell} \] is the background error correlation between levels \( \ell' \) and \( \ell \). The analysis update at level \( l \), \( q_l^A - q_l^B \), is proportional to the inner product of the structure function associated with level \( l \) (row \( l \) of \( B \)) with \( H \), where the constant of proportionality is the factor given in square brackets.

One thing that could go wrong by assimilating TCWV is that the background error covariances are incorrect, resulting in analysis increments wrongly distributed throughout the column. Another is that the observed and modelled scenes have characteristics that are so different that they are incompatible. This difference may arise by, e.g., when the instrument views a dry air column, but the model predicts a moist column at the observation position. This can happen when the background’s boundary layer is too deep (as in Fig. 1a). The resulting negative analysis increment is another means by which moisture analysis may contain negative values. The opposite case, when the background’s boundary layer is too shallow may result in a moisture analysis that is supersaturated, leading to rain-out in the first few model steps. This is the mechanism proposed by [Ingleby et al. (2013)] to possibly explain their excessive precipitation found over the first hour of global forecasts.

### 4.3 Out-of-bounds humidity (negative humidity and unbalanced moisture fields)

The issues mentioned in Sects. 4.1 and 4.2 share the common problematic possibility of negative humidity or supersaturated humidity. Of course, physical principles place bounds on the values that specific and relative humidity can have, i.e. during thermodynamic equilibrium \( 0 \leq q \leq q_{sat} \) and \( 0 \leq \mu \leq 1 \), where \( q_{sat} \) is the specific humidity at saturation. Exceeding the upper bound will create an instability (an imbalance between the moisture and thermodynamic variables), which, as we have stated before, will lead to excessive precipitation early in the forecast following an analysis (spin-down, [Hólm et al. (2002), Ingleby et al. (2013)])

The distribution of three-hour UKV forecasts of \( \mu \) averaged over five days are shown in Fig. 2 for three layers of the model. The BL (panel a) has most points with \( \mu > 0.6 \), has a thin peak at \( \mu = 0.96 \), and relatively few points with \( \mu < 0.3 \). The free troposphere (panel b) shows similar moist features to the BL, but with a broad peak at \( \mu = 0.06 \). The upper troposphere and lower stratosphere (panel c) has a dominant thin peak just above \( \mu = 0 \), indicative of the dry, unsaturated extratropical tropopause. Typical background forecasts of \( \mu \) then have large numbers of points that are close to the natural bounds of humidity. This means that there is a frequent danger that the analysis may tip-over to unphysical values at some points, unless the assimilation has some mechanism of respecting these boundaries. The bottom line is that forecast errors (deviations from particular background values) will have distributions that deviate far from Gaussian.
4.4 The huge variability of specific humidity

Specific humidity varies in the atmosphere by orders of magnitude, and, consequently, so do its forecast errors. Most of the variability is in the vertical direction, over the UK where $q$ forecasts range from $10^{-2}$ kg kg$^{-1}$ at the surface to $10^{-6}$ kg kg$^{-1}$ in the stratosphere. [Dee and da Silva (2003)] raise the issue of assimilating with $q$ as the analysis variable. Using the same conventions as in (5) but for a point observation of specific humidity at level $l'$, $y^O_l$ gives analysis at level $l$:

$$q^A_l = q^B_l + B_{ll'} \left[ \frac{y^O_{ll'} - y^m_{ll'}}{\sigma_{ll'}^O + B_{ll'}} \right] = q^B_l + \sigma^B_{ll'} r_{ll'} \left[ \frac{y^O_{ll'} - y^m_{ll'}}{\sigma_{ll'}^O + \sigma_{ll'}^B} \right]. \quad (6)$$

[Dee and da Silva (2003)] point out that such an observation can detrimentally affect distant levels due to the effect of the background error correlations. For the observation to lead to a modification at level $l$ by more than one background standard deviation there requires $r_{ll'} (y^O_{ll'} - y^m_{ll'}) / \sigma^B_{ll'} > 1 + (\sigma^O_{ll'}/\sigma^B_{ll'})^2$, which can conceivably happen when the innovation is large and the vertical correlations between $l$ and $l'$ have not appreciably reduced from unity. Such a condition applies to Fig. (1)a when $l$ is just above $l'$, but may happen in more general circumstances, especially when the background error statistics are specified too small at the observation position, $\sigma^B_{ll'}$, for this situation.

4.5 Aliasing of cloud and precipitation errors in the radiative transfer model

Satellite radiance measurements are traditionally used only when the scene is clear of precipitation and cloud. Given that many measured MW and IR channels have sensitivity to precipitation and cloud, failing to account for these when they are present can lead to the non-cloud-related retrieved quantities being in error.

[Deblonde and English (2003)] for instance study cloudy and non-cloudy scenes with 1D-Var and synthetic SSMIS data using one retrieval method that allows cloud liquid water to be varied during the procedure (via a variable related to total water, ln$(q + q_l)$), and another method that allows only for the vapour phase ln$q$ (plus a liquid water path variable). Supersaturation is penalised in the method that allows only vapour, and temperature and surface wind speed are also retrieved in both methods. They find that the ability to retrieve temperature is degraded in cloudy scenes when liquid cloud is not considered. Problems like this happen because errors in the radiative transfer calculation due to the inability to control the cloud means that the retrieved temperature (and in general vapour) is changed to incorrect values to compensate [Gong and Holm (2011)]. To gauge the effect that cloud can have on the prediction of the brightness temperature, [Bauer et al. (2011)] show the difference between accounting for and not accounting for cloud in the radiative transfer model (e.g. the differences can be more than 1K in brightness temperature for the AMSU-A channel 4 (52.8GHz), compared to the instrument sensitivity of 0.25K).

There are a number of strategies of dealing with cloud or precipitation affected satellite data apart from discarding it [Bauer et al. (2011)]. One option is to model the effects of cloud and precipitation, not to retrieve cloud and precipitation, but to facilitate a more accurate retrieval of temperature and vapour. This can be done within 4D-Var itself, or via a 1D-Var pre-processing step, e.g. by [Pavelin et al. (2008)] who showed how cloud top pressure and cloud fraction could be extracted within 1D-Var and then passed to 4D-Var for more accurate radiative transfer calculations in cloudy scenes. Another option is to extract cloud and precipitation information from the retrieval for initialisation of models along with temperature and vapour.

4.6 Cloud and precipitation fields in 4D-Var

In both cases discussed at the end of Sect. 4.5, condensate and precipitation fields will need to be available to the radiative transfer model. This may be done by diagnosing cloud and precipitation from a total water-related variable (called ‘prognostic cloud/precipitation’, as in the UKV system, see
Sect. 3.1) or by having explicitly having separate vapour/condensate/precipitation variables (called ‘prognostic cloud/precipitation’). [Gong and Hólm (2011)]. Each option has its merits, but also its dangers. Fig. 3 shows various options for facilitating cloud-affected radiance assimilation in incremental 4D-Var, showing the forward and adjoint steps (a similar means can be used for precipitation-affected radiance assimilation).

There are two options for doing 4D-Var with a total water variable (panels a and b) and both rely on a so-called incrementing operator, e.g. based on [Smith (1990)], which separates vapour ($q$) and condensate ($q_c$) from total water ($q_T$). Panel (a) could be used when micro-physics is not included in the linear model, so that $q_T$ is propagated (or its perturbation) and is split only at times when observations are made. Panel (b) could be used when micro-physics is included, but requires linearisation of the micro-physics to propagate $\delta q$ and $\delta q_c$ separately. The incrementing operator though needs closure assumptions, which will lead to errors when these assumptions are (almost inevitably) not met.

A third option (within dashed line of panel b) uses separate vapour and condensate control variables, and so the incrementing operator is by-passed (avoiding its contribution to errors). This not only requires micro-physics in the linear model, but also their background error covariances to be prescribed. Either of these can give rise to problems. Strong non-linearities in the real system can lead to convergence problems and to ‘contradictory regions’ in phase space which can lead to an incorrect solution being reached. If this is done poorly) and requires a linearisation of the cloud and precipitation processes to be included in the linearised operator in 4D-Var. Furthermore, an ill-suited model of the background errors (including cross correlations) can lead to issues like with any other control variable in data assimilation.

4.7 Representivity errors

The fine scale structure of vapour, cloud and precipitation fields makes them particularly susceptible to representivity errors. Representivity error arises when an instrument measures a feature that is not representable by the input fields to the observation operator and the error statistics of representivity error appears as an extra contribution to the observation error covariance.

4.8 Destruction / creation of cloud without humidity observations

Should a data assimilation system have the capability of adding or removing cloud from the background state without guidance from observations of humidity? Even though the model’s background cloud forecast will be in error, this information is valuable when moisture or cloud observations are absent. Models that use specific humidity and temperature as prognostic variables\(^2\) may lose or gain cloud due to perturbations in temperature even when specific humidity has zero assimilation increments. This is especially possible close to saturation where the increase (decrease) in temperature may evaporate (condense) cloud water. Background error covariances that neglect cross-correlations between vapour and temperature are prone to this problem. This is the reason why the Met Office adopted their relative humidity control parameter (implying cloud conserving cross-correlations) in their early variational assimilation systems (Sect. 5.2). Finite cross-correlations still exist though in severely sub-saturated air, which creates the opposite problem: an anomalous cross-correlation between vapour and temperature (Sect. 4.9).

4.9 Biases

\(^2\)A similar argument stands for mixing ratio and potential temperature.
Figure 3: Flow of information within an incremental variational assimilation system depending on the use of linear micro-physics within 4D-Var. In (a) 4D-Var has a \( q_T \) (total water) control variable and no micro-physics scheme is used (\( q_T \) is propagated by the linear model and \( q \) (vapour) and \( q_C \) (cloud) are diagnosed from \( q_T \) at each time). In (b) 4D-Var has a total water control variable and micro-physics is used in the linear model (\( q \) and \( q_C \) are diagnosed at the beginning, but are propagated separately). The dotted line encloses the case when \( q \) and \( q_C \) are separate control variables and so \( q_T \) is not required. Inc. Op. stands for Incrementing Operator (separating \( q_T \) into \( q \) and \( q_C \)), adj. stand for adjoint, RTM stands for Radiative Transfer Model, \( L \) is radiance, superscripts “B” and “ob” stand for background and observation respectively, \( \mathbf{R} \) is the observation error covariance for radiance observations, and variables with a hat are adjoint variables, i.e. \( \hat{x} = \partial J_{ob} / \partial x \), where \( J_{ob} \) is the radiance observation term of the cost function. Cloud condensate is representative of liquid and solid phases, which themselves be represented separately.


5 Choice of parameter for vapour analysis

As explained in Sect. 3.2, a control parameter is a field whose background errors are taken to be uncorrelated with those of other parameters. A control parameter will have non-zero background error correlations between different positions in space (see Sect. 3.2 to see how this is distinct from a control variable). Choosing a control parameter that represents humidity is complicated by the significant cross-correlations that exist between humidity and other background errors.

There are two standard representations of humidity used by meteorologists, namely the specific and relative humidities, and each have been adopted as moisture control parameters in various variational DA systems. Each choice though has its advantages and disadvantages and so alternatives have also been proposed and used. We discuss these choices, and their variants in this section, but defer discussion of non-Gaussian effects to Sect. 7.

5.1 Specific humidity

Specific humidity (SH), $\delta q$, is perhaps the most basic choice of control parameter for water vapour. It has been used in versions of ECMWF, NCEP, Meteo-France, HIRLAM and CMC analysis systems. Choosing $\delta q$ means that moisture will be analysed independently (in the absence of observations that are related to both moisture and other variables). In equilibrium conditions of low relative humidity (RH), such as in regions of downwelling and in the stratosphere, SH is a conserved tracer. This means that for finite perturbations of the state vector (but sufficiently small perturbations to avoid non-linear effects), $\delta q$ will not be expected to be physically coupled to changes in other variables$^{3}$. Under these conditions SH is arguably a good choice for a moisture control parameter. In conditions that are saturated (or close to being saturated) though the univariate control parameter $\delta q$ would ignore potentially very strong correlations between $\delta q$ and, e.g. $\delta T$, due to physical processes like condensation and evaporation. Failing to account for such strong correlations when they exist can lead to analysis increments that are suboptimal. One consequence of such a system is that it can destroy or create cloud inappropriately. To see this, consider the following relationship between SH ($q$) and RH ($\mu$): $q = \mu q_s(T, p)$, where $q_s$ is the saturated SH which depends on temperature, $T$, and pressure, $p$$^{4}$. Linearisation of this definition provides a relationship between $\mu$, $q$, $T$, and $p$ increments:

$$
\delta q = q_s(T_0, p_0) \delta \mu + \mu_0 \frac{\partial q_s}{\partial T} \left|_0 \right. \delta T + \mu_0 \frac{\partial q_s}{\partial p} \left|_0 \right. \delta p,
$$

where subscript “0” indicates the linearisation state (LS) value, and $\alpha_0$, $\beta_0$ and $\gamma_0$ are shorthand for the LS-dependent factors. In the absence of an explicit cloud variable in a DA system, a simple criterion for the presence of cloud is the condition $\mu \geq 1$, so any change of $\mu$ around unity due to changes in $q$ or $q_s$ may affect the presence of cloud implied by the state vector. From (7) and ignoring pressure (since its effect is relatively small), the change in RH, $\delta \mu$, given $\delta q$ and $\delta T$ is $\delta \mu = (1/\alpha_0) \delta q - (\beta_0/\alpha_0) \delta T$. For illustration let us take the case when a temperature observation is made in a region that is at or just above saturation in the background, and which leads to the analysis increment $\delta T_A > 0$. For no moisture observations and univariate background error statistics, $\delta q_A = 0$, which results in an implied change in RH in the analysis of $\delta \mu_A = -(\beta_0/\alpha_0) \delta T_A$. This is potentially large enough to take the RH below saturation, thus removing cloud from the state vector, which may not be appropriate. Thus if SH is used as a control parameter then it is not appropriate to assume univariate background error statistics, and there is little benefit of using the CVT method described in Sect. 3.2.

In addition to these complicated flow-dependent and multivariate issues surrounding SH, and to non-Gaussian issues that will be covered in Sect. 7, there is a further difficulty with the choice of SH as

$^{3}$This does not guarantee non-correlation though - for instance variations in the strength of downwelling may bring simultaneous variations in $\delta T$ and $\delta \mu$ at a particular position in space, which would appear as a correlation even though they are not affected by each other directly.

$^{4}$In DA work RH is often given as a fractional, rather than as a percentage of the saturated humidity.
as a control parameter. Its background error standard deviations, $\sigma_{Bq}$, change by orders of magnitude with altitude. This means that local increments $O(\sigma_{Bq}(z_1))$ made as a result of an observation at height $z_1$ can be spread by background error correlations to higher levels (e.g. $z_2 > z_1$) and lead to inappropriately large increments ($\gg O(\sigma_{Bq}(z_2))$) there [Dee and da Silva (2003)], and even to an accumulation of water vapour in the dry and poorly observed lower stratosphere [Rabier et al. (1998)].

5.2 Relative humidity

Relative humidity has been used as a control parameter in versions of the Met Office and NCAR analysis systems. In DA work RH is often given as a fractional, rather than as a percentage of the saturated humidity via $q = \mu_0(T, p)$, as given in Sect. 5.1. Obviously RH changes less dramatically than SH does meaning that the previously mentioned problem due to spreading by the covariances with SH will be largely absent with RH. An RH-based control parameter means that $\delta \mu$ would represent moisture in the control vector, even though the analysis increments would still require $\delta q$. Using RH instead of SH has some interesting consequences. Consider the model variables moisture in the control vector, even though the analysis increments would still require $\delta q$.

\[ \begin{pmatrix} \delta p \\ \delta T \\ \delta q \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \gamma_0 & \beta_0 & \alpha_0 \end{pmatrix} \begin{pmatrix} \sigma_{Bp} & 0 & 0 \\ 0 & \sigma_{BT} & 0 \\ 0 & 0 & \sigma_{Bq} \end{pmatrix} \begin{pmatrix} \delta \mu \\ \delta T \chi \\ \delta \mu \chi \end{pmatrix}, \]

(c.f. (8) with $S = I$ and $U_s = I$). The implied B-matrix (Sect. 3.2) would be:

\[ \mathbf{B}_{\text{imp}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \gamma_0 & \beta_0 & \alpha_0 \end{pmatrix} \begin{pmatrix} \sigma_{Bp}^2 & 0 & 0 \\ 0 & \sigma_{BT}^2 & 0 \\ 0 & 0 & \sigma_{Bq}^2 \end{pmatrix} \begin{pmatrix} 1 & 0 & \gamma_0 \\ 0 & 1 & \beta_0 \\ 0 & 0 & \alpha_0 \end{pmatrix} = \begin{pmatrix} \sigma_{Bp}^2 & 0 & \sigma_{Bq}^2 \gamma_0 \\ 0 & \sigma_{BT}^2 & \sigma_{BT}^2 \beta_0 \\ \sigma_{Bq}^2 \gamma_0 & \sigma_{BT} \beta_0 & \sigma_{Bq}^2 \end{pmatrix}, \]

where $\sigma_{Bp}^2$, $\sigma_{BT}^2$, and $\sigma_{Bq}^2$ are the background error variances of pressure, temperature and relative humidity respectively, and $\sigma_{Bq}^2 = \sigma_{Bp}^2 \gamma_0^2 + \sigma_{BT}^2 \beta_0^2 + \sigma_{Bq}^2 \alpha_0^2$ is the background error standard deviation of specific humidity that is implied by using $\delta \mu$, rather than $\delta q$ as a control parameter. The consequences of a relative humidity control parameter may be understood easily by examining the analysis increments, $\delta \mathbf{x}_\Lambda = \mathbf{B}_{\text{imp}} \mathbf{H}^T (\mathbf{HH}_{\text{imp}} \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y} - \mathbf{Hx}_B)$, when (1) specific humidity is not observed and (2) when it is observed.

1. In the case when only $T$ is observed, $y_T$, with error variance $\sigma_{OT}^2$, $\mathbf{H} = (0 \ 1 \ 0)$, and the analysis increment is,

\[ \begin{pmatrix} \delta p_A \\ \delta T_A \\ \delta q_A \end{pmatrix} = \begin{pmatrix} 0 \\ \sigma_{BT}^2 \beta_0 \\ \sigma_{Bq}^2 \end{pmatrix} \frac{y_T - x_{BT}}{\sigma_{BT}^2 + \sigma_{OT}^2}. \]

Substituting these analysis increments into (7) results in $\delta \mu_A = 0$, thus preserving cloud. A similar result is found when only $p$ is observed.

2. In the case when only $q$ is observed, $y_q$, with error variance $\sigma_{Oq}^2$, $\mathbf{H} = (0 \ 0 \ 1)$, and the analysis increment is,

\[ \begin{pmatrix} \delta p_A \\ \delta T_A \\ \delta q_A \end{pmatrix} = \begin{pmatrix} \sigma_{Bp}^2 \gamma_0 \\ \sigma_{BT} \beta_0 \\ \sigma_{Bq}^2 \end{pmatrix} \frac{y_q - x_B}{\sigma_{Bq}^2 + \sigma_{Oq}^2}, \]

which results in a non-zero $\delta \mu_A$.

These properties of a RH-based control parameter mean that SH will be modified by the analysis even when there are no moisture observations (a desirable property in many situations [Lorenc et al. (1996)]).
but in this case this is done in such a way as to keep RH unaffected. This has benefit when background conditions are close to saturation, where small increments, e.g., $\delta T_A$ could inappropriately create or destroy cloud if $\delta q_A$ were unaffected congruently.

This coupling between $\delta T$ and $\delta q$ is still imposed though when the air has a low humidity. In the case when a model has a cool bias in the stratosphere, analysis increments $\delta T_A$ will be systematically positive, which will be accompanied by increments of water vapour, $\beta_0 \delta T_A$, which will also be systematically positive. This will lead to an anomalous accumulation of water vapour in the stratosphere [Dee and da Silva (2003)]. The value of $\beta_0 = \partial q/\partial T$ can be ~0.5 g/kg/K at the surface (compared to a typical mixing ratio there of 7 g/kg), ~0.15 g/kg/K at ~500 hPa (compared to 1.5 g/kg) and 0.005 g/kg/K at the tropopause (compared to 0.002 g/kg). The anomalous coupling between temperature and specific humidity is unlikely to become a problem in the troposphere given humidity observations and moist processes there, but can be a problem in the stratosphere.

In an RH-based scheme, background error correlations between $\delta \mu$ and other variables are ignored. $\delta \mu - \delta T$ correlations for instance have been reported as being larger than $\delta q - \delta T$ correlations [Dee and da Silva (2003)] (although the generality of this answer is not clear, e.g. LORENC finds the opposite). There are non-Gaussian issues with RH errors and non-Gaussianity is dealt with in Sect. 7.

5.3 Pseudo relative humidity

A pseudo relative humidity (PRH) control parameter has been proposed to deal with some of the difficulties found with SH and RH [Dee and da Silva (2003)]. It may be defined as $\mu_{ps} = q/q_{sat}(T_B, p_B)$ (in [Dee and da Silva (2003)] it is defined in terms of mixing ratio rather than specific humidity). The difference with SH is that here the saturated specific humidity is evaluated at the background, and so is constant over the assimilation. This choice is essentially a scaled SH variable, but where the scaling is dependent upon the background flow. This avoids the problem with the huge variability of SH (Sect. 5.1), and the sometimes anomalous coupling between $\delta T$ and $\delta q$ of RH (Sect. 5.2). As the arguments of $q_{sat}$ are constants in an assimilation cycle, a PRH control variable cannot account for physically appropriate coupling between $\delta T$ and $\delta q$.

5.4 Non-linear and adaptive humidity parameter

5.4.1 The form of the transform

On the basis of the work discussed above, the desirable properties of a moisture control parameter are (i) weak correlation with other control parameters, (ii) near Gaussian background error statistics (iii) homogeneity of variance, (iv) conservation of RH in the absence of moisture observations and when conditions are close to saturation but (v) not when far from saturation. Many of these properties may be gained with the new moisture parameter that Ingleby et al. [Ingleby et al. (2013)] have developed over recent years at the Met Office. The non-linear/adaptive (NLA) transform introduces a new control parameter $\delta \mu_{NLA}$ as follows:

$$\delta q = \frac{\alpha_0}{a} \delta \mu_{NLA} + h\beta_0 \delta T.$$  \hspace{1cm} (12)

This transform is based on (7) with $\delta \mu \to \delta \mu_{NLA}/a$, $\delta p \to 0$ (since pressure perturbations are assumed to have negligible effect), and the inclusion of the coefficient $h$. The parameter $\alpha$ is for normalization and $h$ is either a regression or correlation coefficient (see below). In principle these coefficients may be allowed to change during the variational procedure giving the NLA transform its non-linear property. This can partly account for the non-Gaussianity of moisture variables, but this issue is explored further in Sect. 7.3. It is $h$ though that provides the adaptive property discussed here, namely that if $h = 0$ the transform performs like a scaled SH parameter (Sect. 5.1), if $h = 1$ it performs like an RH parameter (Sect. 5.2) and values in between result in mixed properties.

Note that [Ingleby et al. (2013)] uses $\delta \theta$ instead of $\delta T$, and $q_T$ (total water - see Sect. 6.2) instead of $\delta q$, but we keep $\delta T$ and $\delta q$ here for consistency with the rest of this section.
5.4.2 The coefficients \( a \) and \( h \)

In [Ingleby et al. (2013)] \( a \) and \( h \) come from look-up tables which allow them to vary with a reference RH\(^6\) and vertical model level. The tabulated values are determined from training data of error perturbations \( \delta q \) and \( \delta T \). Suppose that \( \delta q_i \), \( \delta T_i \) and \( \delta \mu_{\text{NLA}} \) make up the \( i \)th sample of \( N \) that fit into a particular bin characterised by the RH and vertical level. We ask the question: what value of \( h \) in (12) best predicts the samples \( \delta q_i \) from \( \delta T_i \) and \( \delta \mu_{\text{NLA}} \)? The form of \( h \) that achieves this is the regression coefficient, \( h_i \):

\[
h_i = h_\alpha = \frac{(\delta q \delta T)_\text{td} - (\alpha_0/a) \langle \delta \mu_{\text{NLA}} \delta T \rangle_\text{td}}{\beta_0 \langle \delta^2 T \rangle_\text{td}},
\]

where the “td” subscript stands for training data, the covariance is \( \langle \delta q \delta T \rangle_\text{td} = \frac{1}{N-1} \sum_i \delta q_i \delta T_i \) and the variance is \( \langle \delta^2 T \rangle_\text{td} = \frac{1}{N-1} \sum_i \delta^2 T_i \). Notice that this result does not depend upon the new variable \( \delta \mu_{\text{NLA}} \) because it is assumed to be uncorrelated with other variables in the CVT formulation \((\langle \delta \mu_{\text{NLA}} \delta T \rangle_\text{td} = 0\), Sect. 3.2\). This form of \( h \) will help to ensure that the correlation \( \langle \delta \mu_{\text{NLA}} \delta T \rangle \) over background errors is small in practice. To see this consider the background error covariance \( \langle \delta \mu_{\text{NLA}} \delta T \rangle \). Using (12) and then (13):

\[
\langle \delta \mu_{\text{NLA}} \delta T \rangle = \frac{a}{\alpha_0} \left( \langle \delta q \delta T \rangle - h \beta_0 \langle \delta^2 T \rangle \right) = \frac{a}{\alpha_0} \left( \langle \delta q \delta T \rangle - \frac{(\delta q \delta T)_\text{td}}{\langle \delta^2 T \rangle_\text{td}} \langle \delta^2 T \rangle \right).
\]

As long as the statistics from the training data are statistically representative of background errors then terms on the right hand side of (14) will approximately cancel to give a small value for the covariance.

The regression coefficient is the formally correct object to use for \( h \), but in [Ingleby et al. (2013)] the correlation coefficient, \( h_\epsilon \) is used instead:

\[
h = h_\epsilon = \frac{(\delta q \delta T)_\text{td}}{\sqrt{(\delta^2 q)_\text{td}} \langle \delta^2 T \rangle_\text{td}}.
\]

This correlation coefficient is found to have a more physically sensible range of values than the regression coefficient. For instance, \( h_\epsilon \) was found to be close to zero in the stratosphere (where \( \delta \mu_{\text{NLA}} \) behaves like SH) and increases towards unity for RH values above 0.7-0.8 (where \( \delta \mu_{\text{NLA}} \) behaves more like RH). This is the desired behaviour of \( h \). For lower RH values though \( h_\epsilon \) goes negative and it is not clear how to interpret this physically in terms of the SH-like or RH-like behaviour of the control parameter. In both (13) and (15), \( h_\epsilon \propto \delta q \delta T \) and a negative correlation between \( \delta q \) and \( \delta T \) is sometimes characteristic of downwelling air which brings adiabatically warmed, but dry air downwards (see Sect. 9.1). This effect is not accounted for in the SH, RH or PRH schemes.

5.5 Unbalanced moisture variable (physical and statistical)

5.5.1 A paradigm for balanced and unbalanced variables

The structure of \( U_p \) (the parameter transform - or balance operator - in the CVT introduced in Sect. 3.2) has a lower-block-diagonal form in many VAR systems like those of the Met Office [Lorenc et al. (1996)], the ECMWF [Derber and Bouttier (1999)], Météo-France [Sadiki and Fischer (2005), Fischer et al. (2005)], WRF [Chen et al. (2013), and others. This matrix structure is convenient not only mathematically, but also for the physical interpretation of the control parameters that follows. The following parameter transform may be regarded as a simple paradigm of the concept of unbalanced
variables:
\[
\begin{align*}
\delta x_{pm} &= U_p \delta \chi, \\
\begin{pmatrix}
\delta p \\
\delta q \\
\delta \psi
\end{pmatrix} &= \begin{pmatrix}
I & 0 & 0 \\
L_{p\psi} & I & 0 \\
L_{q\psi} & L_{q\mu} & I
\end{pmatrix}
\begin{pmatrix}
\delta \dot{\psi} \\
\delta \dot{p}_u \\
\delta \dot{q}_u
\end{pmatrix}.
\end{align*}
\]
(16)

The first equation follows from (1) where \(\delta \chi\) is the vector of control parameters (in (1) this is \(\delta \chi = \Sigma U_p \delta \chi\)), and \(S\) in (1) is, for simplicity, assumed here to be the identity matrix. The second equation is an expanded version where \(\delta \psi\), \(\delta p\) and \(\delta q\) are increments of streamfunction, full pressure and full SH respectively; \(\delta \dot{p}_u\) and \(\delta \dot{q}_u\) are the unbalanced pressure and unbalanced SH respectively; and component vectors of \(\delta x_{pm}\) are related to those of \(\delta \chi\) via the balance operators \(L_{p\psi}\), \(L_{q\psi}\) and \(L_{q\mu}\). We use a vector notation here as we are representing fields rather than values at individual locations. Equation (16) represents a hierarchy of parameters where the highest parameter is \(\delta \dot{\psi}\) and the lowest is \(\delta \dot{q}_u\). The adjectives “full”, “balanced” and “unbalanced” used here are important in this paradigm and are now explained with reference to (16).

In (16) each parameter is decomposed in terms of its balanced and unbalanced parts whose sum is the full parameter. The exception though in this scheme is the streamfunction increment, which is assumed to be completely balanced (unbalanced part zero - this makes physical sense in synoptic-scale extra-tropical systems since there the rotational wind is largely geostrophically balanced). In many DA schemes geophysical balance operators are often applied in the CVT, but the terms “balanced” and “unbalanced” have more generalised meanings. The generalised meaning of the “balanced component” is the component of a parameter that is completely related to control parameters lower down in the hierarchy, and the generalised meaning of the “unbalanced component” is the residual, or the part that is uncorrelated with lower parameters.

This system can be seen by expanding out the second and third lines of matrix equation (16). The second line gives \(\delta p = L_{pq} \delta \dot{\psi} + \delta \dot{p}_u\) where \(L_{pq} \delta \dot{\psi}\) is the balanced pressure increment and \(\delta \dot{p}_u\) is the (unbalanced) residual. The operator \(L_{pq}\) may in practice represent a dynamical balance operator like the linear balance equation, or may represent a statistical regression. The third line gives \(\delta q = L_{q\psi} \delta \dot{\psi} + L_{qp} \delta \dot{p}_u + \delta \dot{q}_u\) where \(L_{q\psi} \delta \dot{\psi} + L_{qp} \delta \dot{p}_u\) is the balanced SH increment and \(\delta \dot{q}_u\) is the (unbalanced) residual.

There is a connection between the third lines of Eqs. (7) and (16) (the former in vector notation is \(\delta q = G_0 \delta p + B_0 \delta T + A_0 \delta \mu\)). Assuming here that temperature can be diagnosed from \(\delta p\) via the operator \(\delta T = L_{T\mu} \delta p\), then (7) becomes \(\delta q = A_0 \delta \mu + (B_0 L_{T\mu} + G_0) \delta p\), and using the second line of (16), \(\delta q = A_0 \delta \mu + (B_0 L_{T\mu} + G_0) \delta \dot{p}_u\). This has a similar form to the third line of (16) where \((B_0 L_{T\mu} + G_0) \delta \dot{p}_u = L_{q\psi} \delta \dot{\psi} + \delta \dot{q}_u\). Thus, \(A_0 \delta \mu\) may be interpreted as an unbalanced SH. In a similar way, if the NLA transform (Sect. 5.4) is used, \(h \delta q \delta T\) and \((\alpha_0/\alpha) \delta \mu_{\text{NLA}}\) may be considered to be, respectively, the balanced and unbalanced components of SH.

5.5.2 Generalising the idea of balanced and unbalanced variables

In lower-block-diagonal CVTs the order of the control parameters in the CVT shows the place each parameter has in the hierarchy. In general the first control parameter may be denoted \(\delta \chi^{(1)}\) and the \(i\)th control parameter \((i > 1)\) may be denoted \(\delta \chi^{(i)}_{au}\). The associated PMS variable (Sect. 3.2), \(\delta x^{(i)}_{pm}\), is then:

\[
\delta x^{(i)}_{pm} = L_{u1} \delta \chi^{(1)} + \sum_{j=2}^{i-1} L_{ij} \delta \chi^{(2)}_{au} + \delta \chi^{(i)}_{au}.
\]
(17)

\(^9\)In operational systems there are more variables than those shown in (16), but the simplified form here is sufficient to introduce the idea of unbalanced variables.
Variables may be partitioned into their balanced and unbalanced parts using concepts of dynamical balance, but the idea of balanced and unbalanced variables is more general than this. Generally the term “unbalanced parameter” may be interpreted as a parameter that is assumed to be statistically uncorrelated with parameters lower down the hierarchy. This model of background error covariances allows intuitive and physically meaningful couplings to be represented in DA.

5.5.3 Statistical regressions

Some systems use dynamical balance operators for the mass-wind coupling (denoted in (16) as $L_{p\psi}$). This includes the linear balance equation in the Met Office system [Lorenc et al. (2000)], and the (linearised) non-linear balance equation in the ECMWF [Fisher (2003)] and MM5 [Barker et al. (2004)] systems. Other systems use statistical regressions instead, such as earlier versions of the ECMWF system [Derber and Bouttier (1999)], the Météo-France system [Berre (2000)], and the WRF DA system [Chen et al. (2013)]. The two approaches have pros and cons as follows.

- Dynamical balance operators are based on dynamical theory and those balance equations that depend on the flow give rise naturally to flow-dependent background error statistics. They do however often rely on numerical solvers, which may be expensive. It is often unclear how to use dynamical theory to determine a valid balance equation that diagnoses some variables like moisture.

- Statistical regressions must be determined off-line and flow-dependency has to be added explicitly using (e.g.) look-up tables. It is generally possible to relate any variable to any other using statistical regressions.

It is possible to use dynamical balance operators for some variables and regressions for others. On the basis of the above arguments, the parts of the CVT involving moisture are often modelled with statistical regressions. For instance, in [Chen et al. (2013)] the PMV for moisture is RH and the control parameters are streamfunction ($\delta\psi$), unbalanced velocity potential ($\delta\tilde{\chi}_u$), unbalanced temperature ($\delta T_u$), unbalanced surface pressure ($\delta\tilde{p}_s$), and unbalanced RH ($\delta\tilde{\mu}_u$). The RH part of the parameter CVT may be written as:

$$\delta\mu = L_{\mu\psi}\delta\tilde{\psi} + L_{\mu\chi_u}\delta\tilde{\chi}_u + L_{\mu T_u}\delta T_u + L_{\mu p_s}\delta\tilde{p}_s + \delta\tilde{\mu}_u,$$

where the same convention of labelling the balance (regression) operators is used as in Sect. 5.5.2. Extensions may be made for other variables including hydrometeors. There is a standard regression formula that gives the balance operators $L_{\mu\psi}$, $L_{\mu\chi}$, $L_{\mu T}$, and $L_{\mu p_s}$:

$$
\begin{pmatrix}
L_{\mu\psi} & L_{\mu\chi_u} & L_{\mu T_u} & L_{\mu p_s}
\end{pmatrix}
\begin{pmatrix}
C_{\mu\psi} & C_{\mu\chi_u} & C_{\mu T_u} & C_{\mu p_s}
\end{pmatrix}
= \left( \begin{pmatrix}
C_{\psi\psi} & C_{\psi\chi_u} & C_{\psi T_u} & C_{\psi p_s}
C_{\chi_u\psi} & C_{\chi_u\chi_u} & C_{\chi_u T_u} & C_{\chi_u p_s}
C_{T_u\psi} & C_{T_u\chi_u} & C_{T_u T_u} & C_{T_u p_s}
C_{p_s\psi} & C_{p_s\chi_u} & C_{p_s T_u} & C_{p_s p_s}
\end{pmatrix}
\right)^{-1},
$$

(19)

where $C_{ij}$ is the submatrix of covariances between parameters $i$ and $j$, calculated from training data (individual $C_{ij}$ matrices need not be square, but the full matrix on the right hand side of (19) is square). Note that since inter-parameter correlations are assumed to be zero, the matrix to be inverted is usually assumed to be block-diagonal. In this case, the expression for each $L_{p_1 p_2}$ (where $p_1$ and $p_2$ are each one of the parameters above) is $L_{p_1 p_2} = C_{p_1 p_2} C_{p_2 p_2}^{-1}$ (see analogous expression in [Michel et al. (2011)]).

5.5.4 State-dependent statistics

The implied covariances of model variables are often static. This is a consequence of using a fixed covariance model, e.g. by assuming that the variance statistics and the regression coefficients in (19) are fixed. These assumptions though are not valid in general. For example, the relationship
between RH and temperature may be expected to depend on the closeness to saturation, and on the presence of rainfall. Schemes based around regressions such as those described in Sect. 5.5.3 may gain flow dependence by computing multiple control parameter variances ($\Sigma^2$ in (1)) and multiple sets of regression coefficients for different flows.

[Lorenc (2007)] for instance found that the background error variances of quantities like temperature and RH are highly dependent on the presence of a cloud layer, in particular that variances used in the DA at the top of strato-cumulus layers are only half of their correct values. The static covariance structures are also too strong across cloud-top inversions and the inversions are often forecast in the wrong place [Lorenc (2007), Fowler at al. (2010)]. Reducing covariance lengthscales across inversions may be mitigated with an adaptive mesh method as in [Piccollo and Cullen (2011)], and incorrectly placed inversions can be corrected for in a ‘phase-correcting’ DA scheme such as in [Fowler et al. (2012)] (although ideally the source of such position errors should be found in the forecast models).

There is found to be a clear dependence of background error covariances to cloud, fog and precipitation, which has been dealt in a series of papers using geographical masks which separate regions into precipitating and non-precipitating regions ([Montmerle and Berre (2010), Montmerle (2012)]), different grades of precipitation ([Michel et al. (2011)]) and foggy and clear regions ([Menetrier and Montmerle (2011)]). The core idea in these papers is to construct a CVT of the form:

$$\delta x = \left( F^{1/2}_1 B^{1/2}_1 \ldots F^{1/2}_{N_c} B^{1/2}_{N_c} \right) \begin{pmatrix} \chi_1 \\ \vdots \\ \chi_{N_c} \end{pmatrix}, \quad (20)$$

where $B^{1/2}_i$ is the CVT for the $i$th category (of $N_c$ categories), with control sub-vector $\chi_i$, and $F^{1/2}_i$ is the geographical mask for that category. Examples of categories are heavy, light and no precipitation ($N_c = 3$) or just precipitating and non-precipitating regions ($N_c = 2$). Each $B^{1/2}_i$ (e.g. of the form (17)) is tuned for its particular category. If each $B^{1/2}_i$ is modelled in physical space then each $F^{1/2}_i$ is a diagonal matrix, with values most simply as 1 when the category is relevant for that position, and 0 otherwise, or perhaps more pragmatically as smoothed masks, subject to $\sum_{i=1}^{N_c} (F^{1/2}_i)_{jj} = 1$. The works mentioned above model each $B^{1/2}_i$ in spectral space, where $F^{1/2}_i$ is no longer diagonal, but has the form $F^{1/2}_i = S D^{1/2}_i S^{-1}$, where $S$ is the Fourier transform operator and $D^{1/2}_i$ now takes the role of the mask in physical space.

Some interesting findings from [Montmerle and Berre (2010)] are that precipitating regions show a large increase in $\delta q - \delta \chi$ coupling, a large increase in wind standard deviations, a decrease in $\delta q$ length-scales, and a decrease in $\delta q$ standard deviations compared to non-precipitating regions. These results are perhaps not surprising (the last one may be explained by the background’s $q$ being tied to $q_{sat}$), and they highlight that a flow-dependent formulation like (20) is important. [Montmerle (2012)] in particular showed that such a scheme gave improved forecasts, including better alignment of convective cells (although the innovation in this study also included control variables for cloud and rain content).

Upcoming methods like hybrid En4D-Var merge the static B-matrix of 4D-Var with the completely covariances found from an ensemble [Clayton et al. (2012), Bannister (submitted)]. Such methods naturally allow flow-dependency in the combined B-matrix, although it is arguable that they should not be taken as a substitute for not improving the B-matrix of 4D-Var, or by not allowing it to be flow-dependent itself.
5.6 Variables based on the equations of motion

6 Heterogeneity

6.1 Separate variables for each phase

6.2 Total water variable

6.3 Additional hydrometeors

7 Non-Gaussianity

7.1 Consequences of using a Gaussian control variable for a NG variable

7.2 Log normal distributions (Fletcher)

7.3 Symmetrizing transforms (Holm)

Some aspects of non-Gaussianity of moisture can be accommodated in the usual Var. framework using the procedure reported in [Holm et al. (2002)] in the ECMWF system (and later incorporated by [Ingleby et al. (2013)] in the Met Office system). The essential steps are as follows (note that the description is given here in terms of the analysis variable used by the ECMWF in this work, $\delta \mu$, but it is also applicable to the NLA variable, $\mu_{NLA}$, discussed in Sect. 5.4.

1. Find a representation of humidity perturbations that has the least non-Gaussian characteristics when its PDF is conditioned on similar values of the background RH. Candidates considered in [Holm et al. (2002)] included SH, the logarithm of SH, and RH. The PDFs are found from data (e.g. differences between pairs of forecasts). The variable chosen in [Holm et al. (2002)] is RH, $\delta \mu$, which has conditional background PDF $p(\delta \mu|\mu)$ (where $\mu$ is the 'true' RH for a particular forecast pair, in practice taken to be one of the forecasts in the pair, or their mean as is done below). The conditional PDF is found by binning the difference data according to $\mu$.

2. Even though $p(\delta \mu|\mu)$ is the most Gaussian-like of the candidates, it still has undesirable properties for $\delta \mu$ to be used as a control parameter. Significantly, $p(\delta \mu|\mu)$ significant biases and asymmetry (the latter is found especially when $\mu \sim 0$ and $\mu \sim 1$).

3. A new variable, $\delta \tilde{\mu}$, is introduced, which is defined as $\delta \tilde{\mu} = \delta \mu/\sigma(\mu)$, where for calibration, $\delta \mu$ is the difference between two possible background forecasts ($\delta \mu = \mu^{b_2} - \mu^{b_1}$), and $\sigma(\mu)$ is the background error standard deviation for the bin characterised by the mean of the two forecasts ($\mu = (\mu^{b_2} + \mu^{b_1})/2$). This new variable is useful because, under the condition that the two states $\mu^{b_1}$ and $\mu^{b_2}$ are drawn from the same PDF, the PDF $p(\delta \mu|\tilde{\mu})$ is symmetric and unbiased. This can be shown by swapping $\mu^{b_1}$ and $\mu^{b_2}$ (these come from the same distribution and so the conditional PDF will be unchanged during this swap). Performing this step gives $p(\delta \mu|\tilde{\mu}) = p(\mu^{b_2} - \mu^{b_1}|(\mu^{b_2} + \mu^{b_1})/2) = p(\mu^{b_1} - \mu^{b_2}|(\mu^{b_1} + \mu^{b_2})/2) = p(-\delta \mu|\tilde{\mu})$. Although this does not guarantee Gaussianity, this symmetrised PDF has favourable properties for DA.

4. $\delta \tilde{\mu}$ is taken as the new control parameter, and is related to $\delta \mu$ by the expression given in point 3 above. In the assimilation though, the standard deviation is conditioned, not on the average of two forecasts, but on the average of the background and the current operating point in the variational process (the latter state is $\mu^b + \delta \mu$). The mean state is thus $\tilde{\mu} = \mu^b + \delta \mu/2$. In practice, this method works by using the transform $\delta \tilde{\mu} = \delta \mu/\sigma(\mu^b + \delta \mu/2)$ at each iteration. The assumption made here is that the background and the current operating point are draws from the same PDF. Clearly this is not a strictly valid assumption - for instance the final operating
point is the analysis, and the analysis has a narrower PDF than the background (in addition, the background and analysis states are not mutually independent). The effect of the different distributions of $\mu^b$ and $\mu^b + \delta \mu$ is neglected in Holm’s scheme.

There are a number of issues that the change of variable from $\delta \mu$ to $\delta \tilde{\mu}$ raises. Firstly, the formally correct PDF to use as the prior PDF in DA is $p(\mu)$ (which is often modelled as $p(\delta \mu | \mu^b)$ in incremental Var). In the procedure above this is replaced by $p(\delta \mu | \tilde{\mu})$. This is no longer strictly the background error PDF, even though it might have some reasonable properties. Secondly the part of the CVT which has to be performed at each Var. iteration (written as $\delta \mu = \sigma(\mu^b + \delta \mu/2)\delta \tilde{\mu}$) is an implicit equation, and furthermore changes with each iteration. This introduces non-linearity into the Var. problem. A significant benefit of this scheme though is that increments are likely to be small for $\mu^b \sim 0, 1$ and $\mu^a \sim 0, 1$ (as long as $\sigma(\mu)$ is small for $\mu \sim 0, 1$), thus discouraging negative humidity and super-saturation.

7.4 Gaussian anamorphosis

8 Multi-variate and flow-dependent aspects

8.1 Errors of the day (e.g. from an ensemble)

8.2 A $J_\infty$-like constraint for moisture

9 Candidates for new moisture control variable(s) at convective-scales

Sections 1 to 8 provide the background for us to propose candidates for a set of new moisture control variables for convective-scale data assimilation. The new set of control variables should allow for ...

9.1 Temperature and humidity statistics

The non-linear/adaptive (NLA) transform reviewed in Sect. 5.4 relies on statistics between $\delta q$ and $\beta_0 \delta T$ (in [Ingleby et al. (2013)] the statistics are between $\delta q_T/q_s$ and $\delta \theta(q_T/q_T)\Pi \partial(e_s/\partial T)$). In this section we show regression and correlation coefficients for three bands of RH, for two cases, and for a number of specific model levels.

The signs of the regressions and correlations that we might expect, based on vertical motion (only) in an atmosphere with stratified potential temperature (increasing with height), are summarised in table.

<table>
<thead>
<tr>
<th>Driver process</th>
<th>Consequences</th>
<th>Expected $\delta q$</th>
<th>Expected $\delta T-\delta q$ correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>In updraughts ($w &gt; 0$, assume conditions close to saturation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heating ($\delta T &gt; 0$)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

22
10 Discussion and conclusions

References


[Bannister (in preparation)] Bannister R.N. and co-authors, On dynamical balance conditions in contemporary variational data assimilation applied to high-resolution forecast models (working title), in preparation.


[Bauer et al. (2011)] DOI:10.1002/qj.905


[Bennitt and Jupp (2012)] DOI: 10.1175/MWR-D-11-00156.1


[Dee et al. (2011)] DOI:10.1002/qj.828


[Fisher (2003)] Fisher M., 'Background error covariance modelling'. Pp. 45-64 of ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean, 8–12 September 2003, ECMWF, Reading UK.


[Guan et al. (2011)] doi:10.1029/2010JD015412

[Guidard et al. (2011)] 0.1002/qj.928


[Inglebty (2015)] DOI:10.1002/qj.2372


[Jin et al. (2008)] doi:10.1029/2008JD010040


McNally (2002) DOI:10.1256/qj.01.206


Montmerle et al. (2007) DOI:10.1002/qj.34


Ohring and Bauer (2011) DOI:10.1102/qj.946

Okamoto and Derber (2005) 10.1175/MWR3205.1

Parkinson (2003) 10.1109/TGRS.2002.808319


Pavelin et al. (2008) DOI: 10.1002/qj.243


[Renshaw and Francis (2011)] DOI:10.1002/qj.980


[Sun et al. (2014)] DOI:10.1175/BAMS-D-11-00263.1


[Tong and Xue (2005)] DOI:10.1175/MWR2898.1