Abstract—A variational method to retrieve profiles of temperature and humidity is described, which combines observations from microwave and infrared radiometers and surface sensors with background from short-range Numerical Weather Prediction (NWP) forecasts in an optimal way, accounting for their error characteristics. The required forward models are described. An analysis is presented of the error budget of the background and observations, including radiometric, modeling and representativeness errors. Observation errors of the humidity channels are found to be dominated by representativeness, due to their sensitivity to atmospheric variability on smaller scales than the NWP model grid. Profiles of temperature and total water content are retrieved from synthetic data using Newtonian iteration. An error analysis shows these are expected to improve mesoscale NWP, retrieving temperature and humidity profiles up to 4 km with uncertainties of <1 K and <40% and 2.8 and 1.8 degrees of freedom for signal, respectively, albeit with poor vertical resolution. A cloud classification scheme is introduced to address convergence problems and better constrain the retrievals. This Bayesian retrieval method can be extended to incorporate observations from other instruments and form a basis for future Integrated Observing Systems.

Index Terms—Atmospheric measurements, Microwave radiometry, Remote sensing, Variational methods.

I. INTRODUCTION

Numerical Weather Prediction (NWP) and nowcasting applications have a requirement for observations of temperature and humidity profiles of increasing accuracy, frequency and resolution. It is anticipated that these requirements may be addressed by integrating observations from different ground-based remote sensing instruments, including a microwave radiometer, to supplement the radiosonde network and to complement satellite data over land. These Integrated Profiling Systems offer the potential to provide information on vertical profiles of temperature, humidity and cloud at a high temporal resolution, which could be assimilated into the next generation of convective scale NWP models. This paper demonstrates a one dimensional variational (1D-VAR) retrieval method can be used to combine observations from multiple instruments with background information from an NWP model to provide retrievals of temperature, humidity and cloud profiles. The performance of these retrievals can be compared with the user requirements.

The retrieval of temperature and humidity profiles from passive ground-based sensors is an ill-posed problem, because there are an infinite number of atmospheric states that can produce a given observation vector within its uncertainty. This can be resolved by the addition of background data. Variational retrievals provide an optimal method of combining observations with a background in the form of a short-range forecast from an NWP model, which accounts for the assumed error characteristics of both. For this reason they are often referred to as Optimal Estimation retrievals. The 1D-VAR retrievals presented here are similar to the Integrated Profiling Technique [1], but takes its background from an NWP model instead of radiosondes and uses different control variables to concentrate on retrieving profiles of atmospheric temperature and humidity.

The 1D-VAR retrieval is performed by adjusting the atmospheric state vector, \( \mathbf{x} \), from the background state, \( \mathbf{x}^b \), to minimize a cost function of the form [2]:

\[
J(\mathbf{x}) = \left( \mathbf{x} - \mathbf{x}^b \right)^T \mathbf{B}^{-1} \left( \mathbf{x} - \mathbf{x}^b \right) + \left( \mathbf{y} - H(\mathbf{x}) \right)^T \mathbf{R}^{-1} \left( \mathbf{y} - H(\mathbf{x}) \right)
\]  

where \( \mathbf{B} \) and \( \mathbf{R} \) are the error covariance matrices of the background, \( \mathbf{x}^b \), and observation vector, \( \mathbf{y} \), respectively, \( H(\mathbf{x}) \) is the forward model operator and \( ^T \) and \( ^{-1} \) are the matrix transpose and inverse, respectively, using the standard notation of [3].

II. BACKGROUND DATA AND STATE VECTOR

The mesoscale version of the Met Office Unified Model is used to provide background data for the retrievals in the form of profiles of temperature, humidity and liquid water. The model grid points are interpolated to the position of the observations. This model is initiated every six hours, including data from radiosonde stations. A short-range forecast (T+3 to T+9 hr) is used for the background, as would be available to operational assimilation schemes. This is independent of any radiosondes launched at observation time,
which may be used to validate the retrievals.

The state vector, $\mathbf{x}$, used in the retrievals is defined as the temperature and total water on the lowest 28 model levels. These extend up to 14 km, but are concentrated near the surface, where most of the radiometer’s information is.

In this study the humidity components of the state vector are defined as the natural log of total water, $\ln q$. ($q$ is the specific humidity.) This control variable is a modified version of that suggested in [4], with a smooth transfer function between water vapor for $q < q_{\text{sat}} < 90\%$ and liquid water for $q > q_{\text{sat}} > 110\%$ (where $q_{\text{sat}}$ is $q$ at saturation.) The condensed part of the total water is further partitioned between liquid and ice fractions as a linear function of temperature, producing pure ice at $-40^\circ$C. The choice of total water has the advantages of reducing the dimension of the state vector, enforcing an implicit super-saturation constraint and correlation between humidity and liquid water. The logarithm creates error characteristics that are more closely Gaussian and prevents unphysical retrieval of negative humidity.

The background error covariance, $\mathbf{B}$, describes the expected variance at each level between the forecast and true state vector and the correlations between them. In this work, $\mathbf{B}_{\text{ATOVs}}$ was taken from that used to assimilate data from satellite instruments operationally at the Met Office. $\mathbf{B}$ could also be calculated by the NMC method [5], which estimates $\mathbf{B}_{\text{NMC}}$ as the covariance of the state vectors taken from sequential runs of the forecast model, valid at the same time – e.g. the 6 h forecast and the 12 h forecast from the previous model run, 6 h earlier. For the temperature components of $\mathbf{x}$, $\mathbf{B}_{\text{NMC}}$ has much smaller diagonal terms and correlations than $\mathbf{B}_{\text{ATOVs}}$. However, for humidity, $\mathbf{B}_{\text{NMC}}$ increases with height rapidly. The diagonal components of $\mathbf{B}_{\text{ATOVs}}$ are shown for reference in Fig. 3.

### III. OBSERVATIONS

This study uses observations from the Radiometrics TP/WVP-3000 microwave radiometer [6]. This has 12 channels: seven in the oxygen band 51–59 GHz, which provide information primarily on the temperature profile and five between 22–30 GHz near a water vapor line, which provide humidity and cloud information. This radiometer includes sensors to measure pressure, temperature and humidity at ~1 m above the surface. The pressure is taken as a reference from which geopotential height is calculated at other pressure levels via the hydrostatic equation. The instrument’s integral rain sensor is used to reject periods which may be contaminated by scattering from precipitation, as this is not included in the forward model and emission from raindrops on the radome, which may bias the calibration. This instrument incorporates an optional zenith-viewing infrared radiometer (9.6–11.5 μm) to provide information on the cloud base temperature.

In this study the observation vector, $\mathbf{y}$, is defined as a vector of the zenith brightness temperatures ($T_b$) measured by the radiometer’s 12 channels, with additional elements for the surface temperature ($T_{\text{AMB}}$) and humidity (converted to $\ln q_{\text{AMB}}$) and the infrared brightness temperature ($T_{\text{ir}}$):

$$
\mathbf{y} = [T_{b1}, T_{b2}, \ldots, T_{b12}, T_{\text{AMB}}, \ln q_{\text{AMB}}, T_{\text{ir}}]
$$

The observation error covariance, $\mathbf{R}$, has contributions from the radiometric noise ($\mathbf{E}$), forward model ($\mathbf{F}$) and representativeness ($\mathbf{M}$) errors ($\mathbf{R} = \mathbf{E} + \mathbf{F} + \mathbf{M}$).

The radiometric noise, $\mathbf{E}$, can be evaluated as the covariance of $\mathbf{y}$ measured while viewing a stable scene (such as a liquid nitrogen target) over a short period (~30 min). This term is approximately diagonal – i.e. the channels are independent – with diagonal terms ~0.1–0.2 K$^2$, except the 57.29 GHz channel of this particular instrument, as shown in Table 1.

#### TABLE 1  DIAGNOSTIC COMPONENTS OF OBSERVATIONS ERROR COVARIANCE MATRIX, $\mathbf{R}^{1/2}$, EVALUATED FOR CLEAR AND CLOUDY CONDITIONS

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>22.235 GHz</td>
<td>0.17</td>
<td>0.83</td>
<td>0.65</td>
<td>1.07</td>
</tr>
<tr>
<td>23.035 GHz</td>
<td>0.12</td>
<td>0.84</td>
<td>0.67</td>
<td>1.08</td>
</tr>
<tr>
<td>23.835 GHz</td>
<td>0.11</td>
<td>0.82</td>
<td>0.69</td>
<td>1.08</td>
</tr>
<tr>
<td>26.235 GHz</td>
<td>0.13</td>
<td>0.67</td>
<td>0.78</td>
<td>1.04</td>
</tr>
<tr>
<td>30.000 GHz</td>
<td>0.21</td>
<td>0.61</td>
<td>1.00</td>
<td>1.19</td>
</tr>
<tr>
<td>51.250 GHz</td>
<td>0.18</td>
<td>1.10</td>
<td>1.70</td>
<td>2.04</td>
</tr>
<tr>
<td>52.280 GHz</td>
<td>0.15</td>
<td>0.88</td>
<td>1.35</td>
<td>1.62</td>
</tr>
<tr>
<td>53.850 GHz</td>
<td>0.17</td>
<td>0.35</td>
<td>0.32</td>
<td>0.50</td>
</tr>
<tr>
<td>54.940 GHz</td>
<td>0.18</td>
<td>0.06</td>
<td>0.10</td>
<td>0.14</td>
</tr>
<tr>
<td>56.660 GHz</td>
<td>0.19</td>
<td>0.05</td>
<td>0.10</td>
<td>0.22</td>
</tr>
<tr>
<td>57.290 GHz</td>
<td>0.54</td>
<td>0.05</td>
<td>0.40</td>
<td>0.67</td>
</tr>
<tr>
<td>58.800 GHz</td>
<td>0.18</td>
<td>0.06</td>
<td>0.11</td>
<td>0.22</td>
</tr>
<tr>
<td>$T_{\text{AMB}}$</td>
<td>0.24</td>
<td>0.00</td>
<td>0.15</td>
<td>0.28</td>
</tr>
<tr>
<td>$\ln q_{\text{AMB}}$</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
</tr>
<tr>
<td>$T_{\text{ir}}$</td>
<td>2.50</td>
<td>0.59</td>
<td>8.77</td>
<td>9.14</td>
</tr>
</tbody>
</table>

The forward model error, $\mathbf{F}$, includes contributions from uncertainties in the spectroscopy and errors introduced by the profile discretization, FAP model and monochromatic equivalent frequency (see section IV). The spectroscopic component was estimated as the covariance of the difference in zenith $T_b$ calculated using two absorption codes ([7] and [8]). The other terms were calculated as the covariance of the difference between $T_b$ calculated using the full line-by-line model at high vertical resolution and the approximations. $\mathbf{F}$ contains significant off-diagonal terms, and is largest for the channels most sensitive to the water vapor continuum (26–52 GHz), where it reaches ~0.1 K$^2$.

The representativeness error, $\mathbf{M}$, allows for the radiometer’s sensitivity to fluctuations on smaller scales than can be represented by the NWP model. It is possible to estimate $\mathbf{M}$ by studying the fluctuations in the radiometer’s signal on typical time scales taken for atmospheric changes to advect...
across the horizontal resolution of the NWP model. In the case of the mesoscale model with a 12 km grid, 1200 s was chosen to represent a typical advection timescale. The r.m.s. difference (divided by $\sqrt{2}$) in $y$ measured over this time interval was used to calculate $\mathbf{M}$, after subtracting the contribution from the radiometric noise, $\mathbf{E}$. This showed strong correlation between those channels sensitive to liquid water, water vapor and temperature, respectively. The liquid water and humidity terms were found to vary by an order of magnitude, depending on the atmospheric conditions. The average values calculated over a 7 day period of dry conditions with variable cloud amounts were taken to be typical. This period was later sub-divided into clear and cloudy samples and $\mathbf{M}$ re-evaluated for each. The representativeness term evaluated in this way dominates the observation error covariance of some channels, with terms $\sim(0.1\text{-}1.7 \text{ K})^2$. $\mathbf{M}$ can also be evaluated dynamically, based on time series of observations within 1 hour window of each observation. This technique allows the errors to be reduced in periods of atmospheric stability, when more confidence can be placed that the radiometer observations are representative of the model’s state.

The magnitude of the diagonal components of each term of $\mathbf{R}^{1/2}$ is shown in Table 1 for the 12 channels of the microwave radiometer, surface temperature and humidity sensors (as dimensionless ln$q_t$) and infrared radiometer.

IV. FORWARD MODEL AND ITS JACOBIAN

A forward model, $H(x)$, is needed to transform from state space to observation space. For the microwave radiometer, each channel’s $T_b$ is calculated at an equivalent monochromatic frequency [9] using the radiative transfer equation to integrate down-welling emissions from each atmospheric layer between model levels using a standard absorption model [5], which was found to have small biases in these channels [10]. The forward model for the surface temperature and humidity sensors is trivial – a 1:1 translation to the lowest level of the state vector, $x$. A simple forward model defines $T_r$ as the temperature of the lowest level with any cloud. A more sophisticated radiative transfer model is used here to calculate $T_r$ which accounts for absorption by atmospheric water vapor and the finite extinction in liquid water cloud, assigning extinction coefficients of 33.3 Np/km.(kg/m$^3$)$^{-1}$ and 0.02 Np/km.(kg/kg)$^{-1}$ respectively. This model gives more Gaussian error characteristics, due to having less abrupt transitions at cloud boundaries. Examples of the forward model and its Jacobian are shown in Fig. 1 and Fig. 2.

The Jacobian is the matrix of the sensitivity of the observation vector, $y$, to perturbations of each element of the state vector, $x$, $\mathbf{H}=H'(x)=\nabla_x y$. It is needed to minimize the cost function (see section VI). In this study, $\mathbf{H}$ is calculated by brute force – each level of the state vector, $x$, is perturbed by 1 K in temperature or 0.001 in ln$q_t$. The magnitude of these perturbations was selected to ensure linearity of $\mathbf{H}$, while preventing numerical errors due to truncation.

However, to speed up the calculation, a Fast Absorption Predictor model is used to calculate the absorption in each level below 100 hPa as a third order polynomial function of pressure, temperature and $q$ following [1]. This introduces an additional error in the calculation of $T_b$ described above. $\mathbf{H}$ is only calculated for levels between 0-8 km, corresponding to the maximum range of likely impact from the radiometer data. For levels above this, $\mathbf{H}=0$.
1D-VAR Retrievals

analysis, $A$, is given [2] by:

$$A = \left( H^T R^{-1} H + B^{-1} \right)^{-1}$$

(3)

where $H_i$ is evaluated at the solution (or final iteration).

It is also possible to express the information content of the observations with respect to the background as the Degrees of Freedom for Signal, DFS. This represents the number of layers in the retrieved profile which are retrieved independently. It can be calculated [2] as:

$$DFS = \text{Tr}\left(I - AB^{-1}\right)$$

(4)

where $I$ is the identity matrix and Tr() is the trace operator.

$A$ has been evaluated for different combinations of instruments for a clear US standard atmosphere in Fig. 3, although it depends on the reference state through $H_i$. This shows error in the temperature profile retrieved from the radiometer is expected to approach 0.1 K near the surface, but increases with height, to exceed 1 K above 5 km and includes 2.8 degrees of freedom. For the humidity profile, $A$ varies greatly with $x$. In this example the retrieval’s $\ln q$ error increases from 0.05 (~5%RH) near the surface to 0.4 (~40%RH) by 3 km and includes 1.8 degrees of freedom, increasing by ~1.0 in cloudy conditions. This presents a substantial improvement on the background and the surface sensors alone, which only influence the lowest 500 m, but falls short of the radiosonde’s accuracy above ~1 km for both $T$ and $\ln q$. Fig. 3 also shows the analysis error resulting from the errors assumed in the assimilation of radiosonde data in the Met Office NWP models. Their $R$ is a diagonal matrix, dominated by representativeness errors and may not be perfect. However, the radiometer provides much more frequent observations than radiosondes can, reducing errors of representativeness applying their data to analysis at arbitrary times.

When $B_{BATS}$ is used instead of $B_{ATOV}$, $A$ gives more limited improvements up to 3 km, providing 2.6 degrees of freedom for temperature, but only 1.2 for humidity – fractionally more than a surface sensor.

Plotted as square root of the diagonal components for the lowest 5km of temperature [K] and humidity ($\ln q$) [dimensionless].

However, $A$ only tells part of the story. The other important aspect of the retrieval’s performance is the vertical resolution – its ability to resolve a perturbation in state space. One simple, robust definition of the vertical resolution is the inverse of the trace of the averaging kernel matrix [2]. This is evaluated in Fig. 4, which shows that the vertical resolution of temperature profiles increases with height, from ~1 km near the surface, as approximately twice the height from 0.5-4 km. For $\ln q$, it increases from ~1.5 km near the surface, as approximately 4 times the height above 1.5 km, but this is critically dependent on the reference state, $x$, due to non-linearity in $H$. However, this definition tends to over-estimate the vertical resolution by a factor of ~2 compared to other methods [11], [12].

Fig. 4. Vertical Resolution of temperature and humidity ($\ln q$) retrievals.

VI. MINIMIZATION OF COST FUNCTION

Variational retrievals are performed by selecting the state vector that minimizes a cost function in the form of (1). For linear problems, where $H$ is independent of $x$, this can be solved analytically. However, the retrieval of temperature profiles above ~1 km and humidity profiles is moderately non-linear, so the minimization must be conducted numerically. This has been achieved using the Levenberg-Marquardt method [2] (which was found to improve the convergence rate in cloudy conditions compared to the classic Gauss-Newton method) by applying the following analysis increments iteratively:

$$x_{i+1} = x_i + \left( (1 + \gamma) B^{-1} + H_i^T R^{-1} H_i \right)^{-1} \times \left[ H_i^T R^{-1} (y - H(x_i)) - B^{-1} (x_i - x^b) \right]$$

(5)

where $x_i$ and $x_{i+1}$ are the state vector before and after iteration $i$, and $H_i$ is the Jacobian matrix at iteration $i$.

This is iterated until the following convergence criteria [2] is satisfied, based on a $\chi^2$ test of the residuals of $[y-H(x)]$:
\[
\left[ (H(x_{i+1}) - H(x_i)) \right]^T S_{yy}^{-1} \left[ (H(x_{i+1}) - H(x_i)) \right] << m
\]

where \( S_{yy} \) is the covariance matrix between \( y \) and \( H(x) \) and \( m \) is the dimension of \( y \) (\( m=15 \) in this case).

This typically takes 3-10 iterations, each requiring \( \sim 0.25 \) s of CPU time on a 2.4 GHz Pentium IV using the Fast Absorption Predictor model.

Upon convergence the retrieved state vector, \( \hat{x} \), is tested for statistical consistency with \( y \) and \( R \) by calculating the value:

\[
\chi^2 = [H(\hat{x}) - y]^T R^{-1} [H(\hat{x}) - y]
\]  

(7)

Retrievals with a \( \chi^2>100 \) were rejected. The choice of \( \chi^2 \) threshold was found not to be critical, as it had a small influence on the statistics of the retrievals.

VII. EXAMPLE 1D-VAR RETRIEVALS

Fig. 5 shows an example of a 1D-VAR retrievals using synthetic observations, generated to be consistent with \( R \). These are based on a real radiosonde profile from Camborne (UK) at 11:21 on 9/12/2004 and NWP background profile from a 5 hr forecast, valid 21 minutes earlier. This case was selected because the model had forecast the inversion \( \sim 200 \) m too low and underestimated the humidity by a factor of \( \sim 2 \) over the whole profile. The retrieval was repeated for 100 such sets of observations, all of which converged within 4 iterations on average. The retrieved profiles are closely clustered with typical standard deviations of 0.2-0.4 K in temperature and 0.05-0.10 in \( \ln q \), showing they are relatively robust in the presence of observation noise. In all cases, the retrieval thins the cloud and gives profiles closer to the truth than the background. However, the correlation between temperature at adjacent levels of \( B \) makes it impossible for the retrieval to move a misplaced feature in the vertical.

VIII. CLOUD CLASSIFICATION SCHEME

Examination of the performance of the retrieval scheme showed there were often problems when the humidity approaches the threshold of cloud formation – the residuals often oscillate without reaching convergence. This was partially improved by the implementation of the Levenberg-Marquardt method of minimization, which adjusts the size of the increment at each iteration to change from the classic Gauss-Newton method towards the method of steepest decent, according whether the previous iteration has reduced the cost function, \( J \).

Convergence problems where \( \ln q \) approaches the cloud threshold can also be caused by the error characteristics of \( T_r \), which can be highly non-Gaussian. This has been addressed by introducing a cloud classification as a pre-processing step to the retrieval, based on a threshold of the infrared brightness temperature, \( T_r \). If the observed (or synthetic) \( T_r > \min \{T_{AMB} - 40 \text{ K}, 223 \text{ K} \} \), the profile is classified as cloudy and the retrieval proceeds as described above. Otherwise, the profile is classified as clear and the control variable changed from \( \ln q \) to the log of the specific humidity, \( q \) and an addition term [13] is added to the cost function to prevent saturation. In clear cases, the representativeness term can be reduced by re-evaluating it in only clear sky conditions to allow more accurate retrievals in clear conditions. Rainy observations are rejected.

IX. STATISTICS OF 1D-VAR RETRIEVALS

1D-VAR retrievals were performed on an extended dataset of 1 year of radiosonde profiles from Camborne, using synthetically generated observations and backgrounds, consistent with \( R \) and \( B \), respectively. The statistics for the combined clear and cloudy cases, shown in Fig. 6, are in good agreement with the expected performance from the error analysis, albeit with a poor convergence rate (\( \sim 50\% \)). There is no significant difference in the performance in clear and cloudy cases. Although the background profiles have a small bias, this is corrected in the retrievals.
The retrieved values of Integrated Water Vapor (IWV) were also compared to the radiosonde values. These were found to be good, with a small bias and a standard deviation of 0.88 kg/m$^2$ (compared to the corresponding value for synthetic backgrounds, 2.00 kg/m$^2$). This compares favorably with other methods, which have been shown to retrieve IWV from microwave radiometer observations with an accuracy of better than 1.0 kg/m$^2$ compared to radiosondes in mid-latitude winter [14]. This implies that the retrievals do not need an additional constraint in the cost function to force the IWV to match that retrieved by a simpler method as this is achieved implicitly in the 1D-VAR retrievals.

X. CONCLUSIONS AND FUTURE WORK

A 1-D variational retrieval has been developed to allow observations from ground-based microwave and infrared radiometers and surface sensors to be combined with a background from an NWP model in an optimal way, which accounts for their error characteristics. This has been used to retrieve profiles of temperature, humidity and cloud using a novel total water control variable. This has been shown to be advantageous over methods taking their background from statistical climatology [15].

The observation errors for channels sensitive to cloud were dominated by representativeness errors. To reduce their impact, these can be evaluated dynamically. Convergence problems were encountered in cloudy cases, partially due to the non-Gaussian error characteristics of the infrared observations. A cloud classification scheme has been introduced to address this and help constrain the retrievals.

The 1D-VAR retrievals also have the advantage of providing an estimate of the error on the retrieved profile. Error analysis has shown the microwave radiometer improves the NWP background up to 4 km, retrieving temperature profiles with <1 K uncertainty and 2.8 degrees of freedom for signal and humidity with <40% uncertainty and 1.8 degrees of freedom. These results depend on the background error covariance. However, the vertical resolution of the retrieved profiles is poor and degrades with height. Furthermore, the retrievals were not able to move a misplaced feature in the background profile.

The variational method allows different instruments to be combined if their observations’ forward model operator and error estimates are available. This provides a basis for the development of Integrated Observing Systems. In the future the 1D-VAR retrievals will be extended to include observations from other instruments, such as the cloud base height from a ceilometer, cloud base/top from cloud radar and refractive index gradient from a wind profiler.

Assimilation of these observations could improve mesoscale Numerical Weather Prediction (NWP), especially boundary layer and cloud properties. However, to fully exploit the high time resolution available from ground-based instruments will require 4-Dimensional Variational Assimilation (4D-VAR).

REFERENCES