CAMS Service Evolution

D2.6 Report on weak-constraint 4DVar

Due date of deliverable	31 December 2025
Submission date	18 December 2025
File Name	CAMEO-D2-6-V1.0
Work Package /Task	2/2.4
Organisation Responsible of Deliverable	ECMWF
Author name(s)	Noureddine Semane
Revision number	V1.0
Status	ISSUED
Dissemination Level	PUBLIC

Funded by the European Union

The CAMEO project (grant agreement No 101082125) is funded by the European Union.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

1 Executive Summary

In WP2 of the CAMEO project, task 2.4 aims to improve the IFS data assimilation (DA) method by incorporating model errors through weak-constraint 4D-Var (WC-4DVar). During the first 18 months of CAMEO, WC-4DVar was extended to stratospheric ozone (see CAMEO deliverable D2.5). Over the past 18 months, we have i) successfully implemented stratospheric humidity analysis in IFS-NWP (to be featured in the upcoming operational cycle 50r1), ii) added water vapour as a control variable in IFS-COMPO (implemented in cycle 50r1), and assimilated EOS-Aura Microwave Limb Sounder (MLS) water vapour data during the Hunga Tonga eruption with promising results, and iii) extended WC-4DVar to humidity. An initial assessment of the WC-4DVar analysis, including humidity, indicates a positive impact and significant operational potential.

Table of Contents

2 Introduction

2.1 Background

Variational data assimilation 4DVar (time plus three spatial dimensions) aims to find a model trajectory that best fits observations over an assimilation time window by adjusting the initial conditions for forward model integration using a least-squares approach. In strong-constraint 4DVar, it is assumed that the forward model perfectly represents the evolution of the actual atmosphere, and the best-fitting model trajectory is obtained by adjusting only the initial conditions through minimisation of a cost function, subject to the model equations as a strong constraint. Conversely, relaxing the assumption that the model is perfect leads to the weak-constraint 4DVar (WC-4DVar) formulation, in which model errors are incorporated as corrections to the time derivatives of model variables. The optimal model trajectory is then found by simultaneously adjusting both the model error and initial conditions (Fisher et al. 2005, Trémolet, 2006, 2007).

WP2 of the CAMEO project, task 2.4, aims to improve the DA methodology by including model errors and dynamical constraints through weak-constraint 4D-Var. Building on the proven effectiveness of WC-4DVar in correcting systematic errors in temperature, divergence, and vorticity in the stratosphere (Laloyaux et al., 2020a, 2020b), this document reports on extending the approach to include humidity throughout the entire atmospheric column complementing the work carried out in the first 18 months of CAMEO to extend WC-4DVar to stratospheric ozone (see CAMEO deliverable D2.5).

2.2 Scope of this deliverable

2.2.1 Objectives of this deliverable

This deliverable describes the work carried out under WP2 Task 2.4 in the last 18 months of the project.

2.2.2 Work performed in this deliverable

In this deliverable the work as planned in the Description of Action (DoA, WP2 T2.4) was performed.

2.2.3 Deviations and counter measures

No deviations have been encountered.

2.2.4 CAMEO Project Partners:

ECMWF	EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS
Met Norway	METEOROLOGISK INSTITUTT
BSC	BARCELONA SUPERCOMPUTING CENTER-CENTRO NACIONAL DE SUPERCOMPUTACION

KNMI	KONINKLIJK NEDERLANDS METEOROLOGISCH INSTITUUT-KNMi
SMHI	SVERIGES METEOROLOGISKA OCH HYDROLOGISKA INSTITUT
BIRA-IASB	INSTITUT ROYAL D'AERONOMIE SPATIALE DE BELGIQUE
HYGEOS	HYGEOS SARL
FMI	ILMATIETEEN LAITOS
DLR	DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV
ARMINES	ASSOCIATION POUR LA RECHERCHE ET LE DEVELOPPEMENT DES METHODES ET PROCESSUS INDUSTRIELS
CNRS	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS
GRASP-SAS	GENERALIZED RETRIEVAL OF ATMOSPHERE AND SURFACE PROPERTIES EN ABREGE GRASP
CU	UNIVERZITA KARLOVA
CEA	COMMISSARIAT A L ENERGIE ATOMIQUE ET AUX ENERGIES ALTERNATIVES
MF	METEO-FRANCE
TNO	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO
INERIS	INSTITUT NATIONAL DE L ENVIRONNEMENT INDUSTRIEL ET DES RISQUES - INERIS
IOS-PIB	INSTYTUT OCHRONY SRODOWISKA - PANSTWOWY INSTYTUT BADAWCZY
FZJ	FORSCHUNGSZENTRUM JULICH GMBH
AU	AARHUS UNIVERSITET
ENEA	AGENZIA NAZIONALE PER LE NUOVE TECNOLOGIE, L'ENERGIA E LO SVILUPPO ECONOMICO SOSTENIBILE

3 4DVAR setup

3.1 Weak constraint 4DVAR formulation

Let the vector $\boldsymbol{x}_k$ be used to represent the state of the atmosphere at the time k, then its evolution accounting for the model error is written as,

$$\boldsymbol{x}_{k+1} = M(\boldsymbol{x}_k) + \boldsymbol{\eta}$$

Where M represents the model and η its error. In this implementation, the model error tendencies are assumed to be constant throughout the entire assimilation window.

The WC-4DVar cost function is given by:

$$\mathrm{J}(\boldsymbol{x}_0, \boldsymbol{\eta}) = \frac{1}{2}(\boldsymbol{x}_0 - \boldsymbol{x}_0^b)^T \mathbf{B}^{-1}(\boldsymbol{x}_0 - \boldsymbol{x}_0^b) + \frac{1}{2}\sum_{k=0}^{N}(H(\boldsymbol{x}_k) - \boldsymbol{y}_k)^T \mathbf{R}_k^{-1}(H(\boldsymbol{x}_k) - \boldsymbol{y}_{\boldsymbol{k}}) + \frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\eta}^b)^T \mathbf{Q}^{-1}(\boldsymbol{\eta} - \boldsymbol{\eta}^b) \quad (2)$$

Where $\boldsymbol{\eta}^b$ is the prior estimate of the model error forcing (which are the model error tendencies) estimated in the previous WC-4DVar analysis update and $\mathrm{Q} = \mathrm{E}[\boldsymbol{\eta}\boldsymbol{\eta}^T]$ is the model error covariance matrix, also called the Q matrix, where E represents the expected value. Comparing the strong (Figure 1a) and weak constraint (Figure 1b), in the formulation of the former, it is assumed that $\boldsymbol{\eta} = \boldsymbol{\eta}^b = 0$. Humidity assimilation relies on prior background constraints (i.e., the short-range humidity forecast and its error covariance) to control the vertical distribution of humidity information.

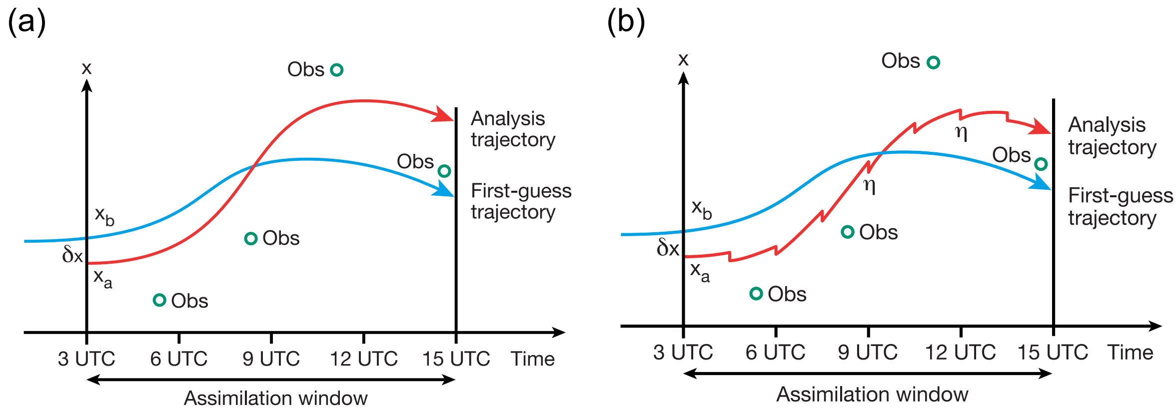

Figure 1: Strong (a) and weak constraint (b) 4DVar.

3.2 ECMWF 4D-Var humidity data assimilation

As described by Bonavita et al. (2016), the IFS 4DVar assimilation system uses a pseudo-relative humidity control variable, defined as specific humidity scaled by the background saturation specific humidity. The main advantage of this approach is that the error statistics for this control variable are like those of a Gaussian distribution, making them more suitable for 4DVar. The latter updates the pseudo-relative humidity using various observations, including radiosonde humidity profiles, GNSS-RO bending angles, and satellite microwave and infrared sounder radiances. The background error covariances, derived from the Ensemble of Data Assimilations (EDA), provide sharper and more localised error estimates. As outlined in Semane and Bonavita (2025), we have successfully reintroduced the analysis of stratospheric humidity into the ECMWF operational assimilation cycle. This will be a key CAMEO contribution to the upcoming IFS Cycle 50r1. While it is not necessary to repeat the complete details presented in Semane and Bonavita (2025), some context on the ECMWF humidity analysis framework and the motivation for the work in this deliverable is helpful. In the current ECMWF Integrated Forecasting System (IFS) data assimilation framework, stratospheric humidity is not analysed using observations. Following an early implementation in 1999 that allowed humidity increments above the tropopause, significant systematic forecast errors were observed. As a result, stratospheric humidity analysis increments were disabled, and the humidity field above the tropopause was effectively taken from the short-range forecast. This left stratospheric moisture largely unconstrained, contributing to known moist and cold biases in lower-stratospheric forecasts. Re-introducing a stratospheric humidity analysis was necessary because new developments, such as ensemble-based background error estimation via the Ensemble of Data Assimilations (EDA), higher vertical resolution near the tropopause, and more sophisticated control variables, now make it feasible to produce physically consistent, observation-informed stratospheric humidity increments that reduce systematic forecast errors. Without first restoring this capability, it would not have been possible to proceed directly to a weak-constraint formulation for humidity, since the infrastructure and error representation needed for such a formulation depend on a constrained analysis framework. This preparatory work constitutes core CAMEO work.

In data assimilation, a control variable is a quantity that the system directly adjusts to minimise the difference between model forecasts and observations. As part of CAMEO, we added CAMS chemical water vapour, the mass of water vapour per unit mass of dry air, as a new control variable. The technical implementation required integration into the IFS assimilation infrastructure, including representation in the background error covariance. This development was essential because, without a properly defined control variable for chemical water vapour, it would not have been possible to assimilate stratospheric water vapour observations to correct the chemical water vapour field. By implementing chemical water vapour as a control variable in IFS Cycles 49r2 and 50r1, we enabled the assimilation of EOS-Aura MLS water vapour retrievals (Semane et al., 2025).

Figure 2 illustrates water vapour latitude-pressure cross sections on 15 January 2022 from BIRA reanalysis[1] as described in Errera et al. (2019) (a), CAMS analysis without MLS water vapour assimilation (b), and CAMS analysis with MLS water vapour assimilation (c). It also shows water vapour latitude-pressure cross sections on 30 June 2022 from BIRA reanalysis (d), CAMS analysis without MLS assimilation (e), and CAMS analysis with MLS assimilation (f). These figures highlight the influence of MLS water vapour assimilation before and six months after the Hunga-Tonga eruption. The CAMS analysis incorporating MLS water vapour data consistently detects the Hunga-Tonga signal, aligning closely with the BIRA reanalysis.

[1] BASCOE Reanalysis of Aura-MLS, version 3 (BRAM3)
Dataset link: https://doi.org/10.18758/8klj122q

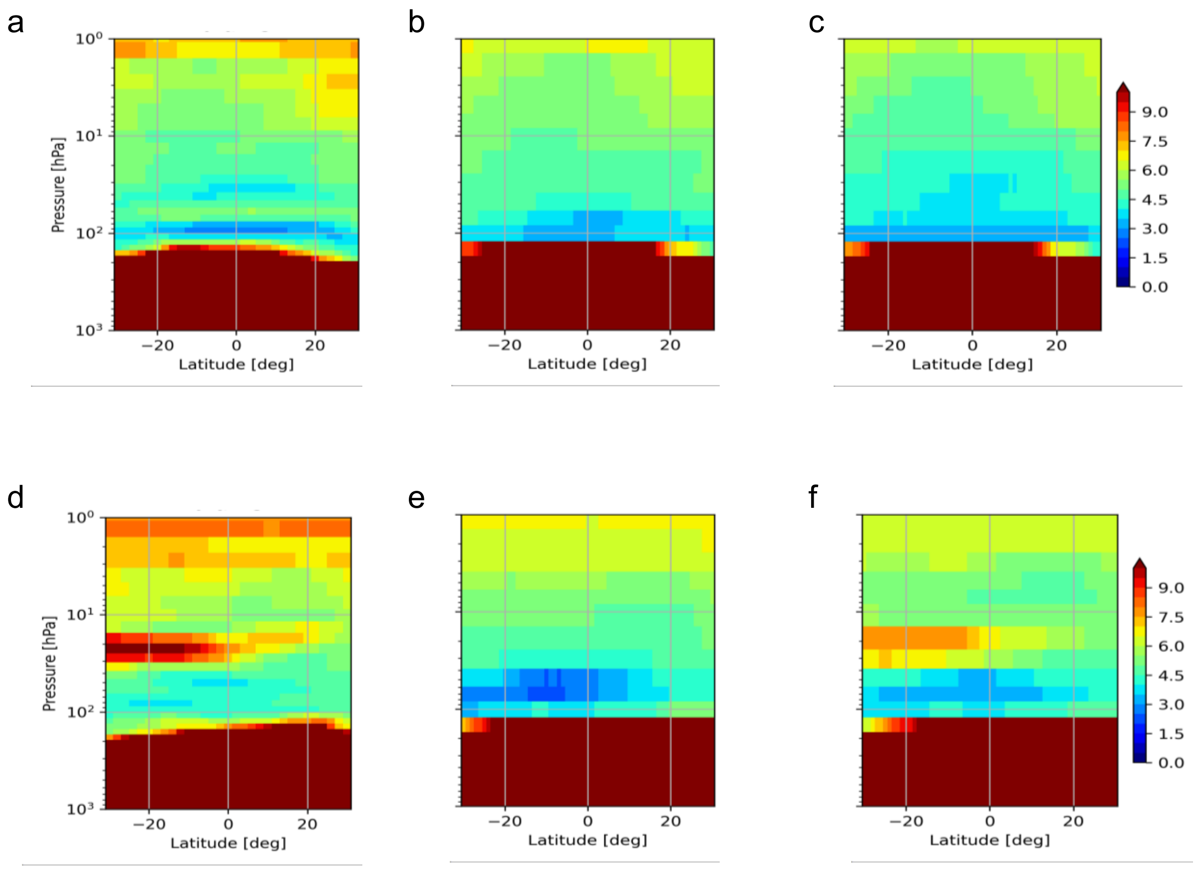

Figure 2: Water vapour latitude-pressure cross section on 15 January 2022 from BIRA reanalysis (a), CAMS analysis without MLS assimilation (b), and CAMS analysis with MLS assimilation (c). Water vapour latitude-pressure cross section on 30 June 2022 from BIRA reanalysis (d), CAMS analysis without MLS assimilation (e), and CAMS analysis with MLS assimilation (f). Those plots are based on an evaluation carried out by Marc Op de beeck and Quentin Errera (BIRA-IASB).

3.3 Neural Network derived model error covariance

Following Bonavita and Laloyaux, 2020, we trained an artificial neural network (ANN) to learn accumulated specific humidity model errors over a 12-hour assimilation window using analysis increments (Analysis-Background) as predictors and a combination of climatological (latitude, longitude, time of day, month) and state-dependent (columns of first-guess forecast fields) predictors. The training dataset for analysis increments and background forecasts is collected over the entire year of 2021 using a 49r1-like analysis experiment with stratospheric humidity analysis turned on. The trained ANN is then employed to generate a representative sample of model errors and derive the model error covariance matrix (Q). The latter is horizontally localised with a cosine function tapering correlations to zero between 4000 and 6000 km, thereby removing spurious hemispheric-wide correlations. Vertically, it is localised with a quadratic function of the distance from the diagonal to control sampling noise.

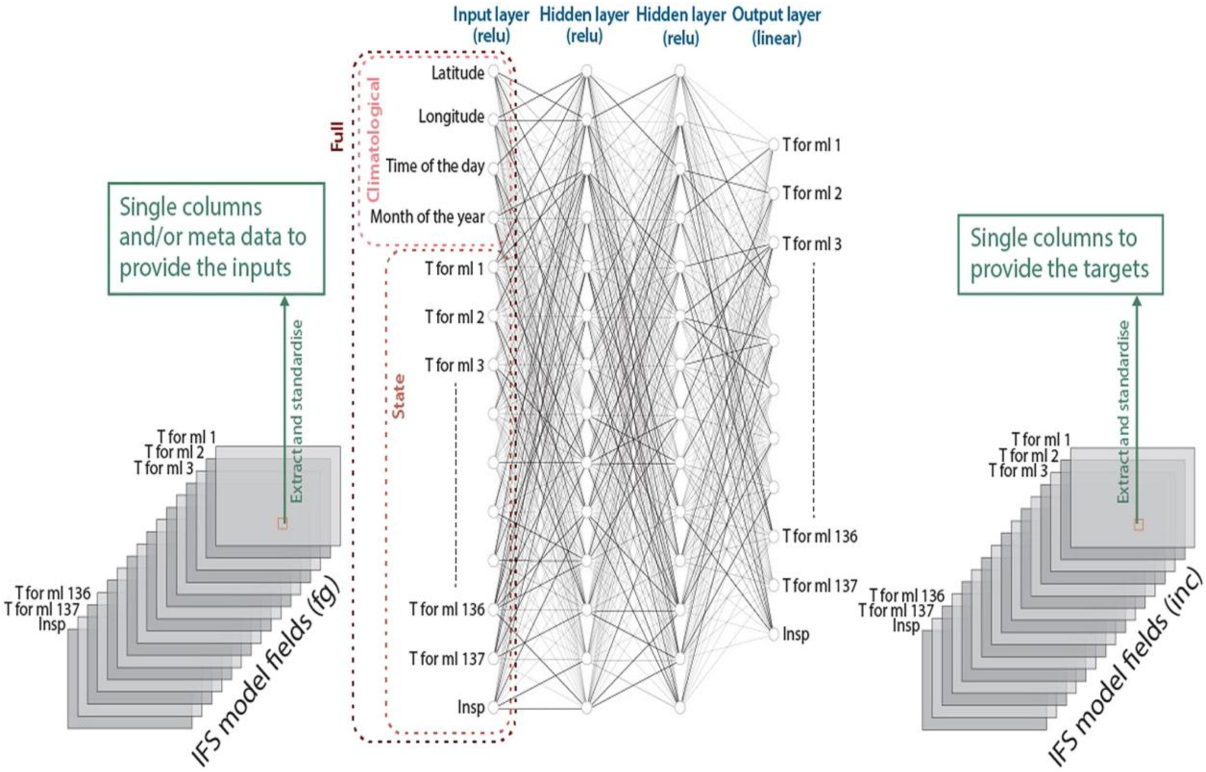

Figure 3: Diagram representing how ANN is built for the regression. Single columns plus metadata (latitude, longitude, time of the day, and month of the year) are extracted from the first guess and analysis increment gridded fields to produce the input and the target of the neural network (Bonavita and Laloyaux, 2020).

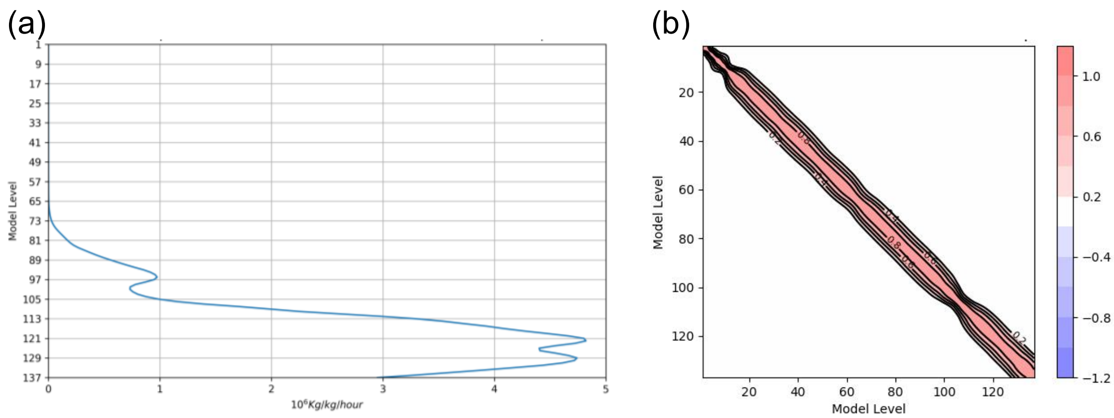

Figure 4: (a) Vertical profile of the ANN-derived standard deviation of model error tendencies for specific humidity. (b) Vertical correlations of ANN-derived model error tendencies for specific humidity. Model level 1 is the top of the atmosphere; model level 137 corresponds to the surface.

Figure 4a displays vertical profiles of the standard deviation of ANN-derived model error tendencies. The vertical structure of model-error correlations shown in Figure 4b is a vital aspect of the Q matrix, as it influences how model-error information is distributed within the atmospheric column.

4 WC-4DVar results

To assess the impact of the weak constraint for specific humidity on the data assimilation system, analysis experiments were conducted over 25 days with (WC-4DVar) and without (CTL) the weak constraint for humidity.

4.1 DA departure statistics

Analysis and background (i.e., the short-range forecast) departure statistics (shown below) from a cycling NWP data assimilation experiment (using the same observations as in operations) are essential tools for verifying the effectiveness of any upgrade to the data assimilation system. Figure 5 shows a significant reduction in analysis and background-observation departures when WC-4DVar is extended to humidity.

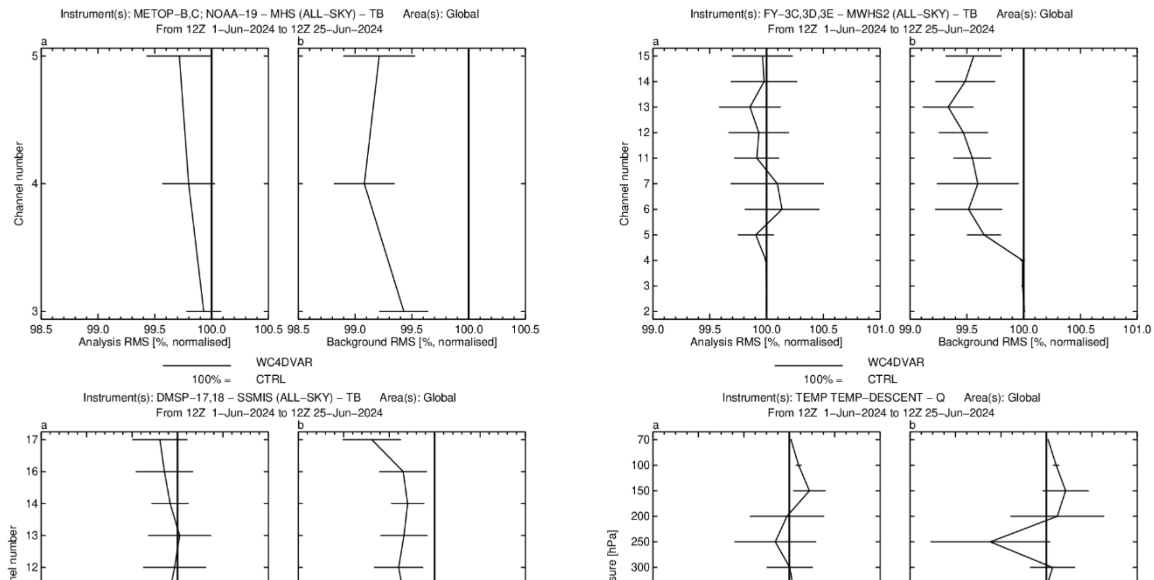

Figure 5: Relative change in the rms analysis and background fits for humidity observations (MHS, MWHS2, SSMI and TEMP-Q). Values lower than 100% indicate that the WC4DVar extended to humidity has smaller analysis and background departures than the control. Horizontal lines indicate the uncertainly margin.

5 Conclusion

In WP2 of the CAMEO project, task 2.4 aims to improve the DA methodology by integrating model errors through weak-constraint 4DVar. Building on the success of WC-4DVar in correcting systematic model errors in stratospheric temperature, divergence, and vorticity, this report discusses its extension to include humidity. After extending WC-4DVar to stratospheric ozone in the first half of CAMEO, we have successfully implemented WC-4DVAR for humidity over the past 18 months of the CAMEO project. Model bias correction for humidity shows that the adjusted first-guess trajectory aligns more closely with the assimilated humidity observations compared to a control experiment without WC-4DVAR. As WC-4DVAR for humidity progresses towards operational deployment, further verification over an extended period is ongoing before potential implementation in the next ECMWF model cycle, CY51R1, or possibly in an intermediate cycle between 50r1 and 51r1.

In IFS-COMPO, water vapour is now implemented as a control variable (Cycles 49r2 and 50r1), and the assimilation of MLS water vapour has been successfully demonstrated in research experiments, providing a strong foundation. However, WC-4DVar has not yet been extended to chemical water vapour. Additional work is needed before applying the same approach used for specific humidity: estimating the model error covariance matrix for water vapour based on ANN-derived model error tendencies for chemical water vapour.

References

Bonavita, M., E. Hólm, L. Isaksen & M. Fisher, 2016: The evolution of the ECMWF hybrid data assimilation system. Q.J.R. Meteorol. Soc., 142, 287-303. https://doi.org/10.1002/qj.2652

Bonavita, M. and Laloyaux, P., 2020: Machine learning for model error inference and correction. Journal of Advances in Modeling Earth Systems, 12, e2020MS002232. https://doi.org/10.1029/2020MS002232

Errera, Q., Chabrillat, S., Christophe, Y., Debosscher, J., Hubert, D., Lahoz, W., Santee, M. L., Shiotani, M., Skachko, S., von Clarmann, T. and Walker, K., 2029: Technical note: Reanalysis of Aura MLS chemical observations, Atmospheric Chemistry and Physics, 19, 21, 13647-13679. https://doi.org/10.5194/acp-19-13647-2019

Fisher, M, Leutbecher, M, and Kelly, G., 2005: On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, Q. J. R. Meteorol. Soc., 131, 3235–3246.

Laloyaux, P, Bonavita, M, Dahoui, M, et al., 2020a: Towards an unbiased stratospheric analysis, Q J R Meteorol Soc, 146: 2392– 2409. https://doi.org/10.1002/qj.3798

Laloyaux, P., Bonavita, M., Chrust, M., and Gürol, S., 2020b: Exploring the potential and limitations of weak- constraint 4D-Var, Q J R Meteorol Soc., 146: 4067– 4082. https://doi.org/10.1002/qj.3891

Semane, N. and Bonavita, M., 2025: Reintroducing the analysis of humidity in the stratosphere, ECMWF Newsletter No. 183 – Spring 2025, pp. 28–31, doi: 10.21957/ns41h80kl3

Semane et al., 2025, Assimilation of Water Vapour in the Upper Troposphere–Lower Stratosphere. In preparation for submission to QJRMS,

Trémolet, Y., 2006: Accounting for an imperfect model in 4D-Var, Q. J. R. Meteorol. Soc., 132, 2483–2504.

Trémolet, Y., 2007: Model error estimation in 4D-Var, Q. J. R. Meteorol. Soc., 133, 1267–1280.

Document History

Version	Author(s)	Date	Changes
0.1	Noureddine Semane	5 December 2025	Initial version
1.0	Noureddine Semane	Dec 2025	Updated after internal review

Internal Review History

Internal Reviewers	Date	Comments
Zoi Paschalidi (ECMWF) and Marion Schroedter-Homscheidt (DLR)	Dec 2025	