Preliminary Assessment of MetOp-Based Temperature and Humidity Statistical Retrievals within the 3D-Var AROME-France Prediction System

Bruna Barbosa Silveira, Nadia Fourrié, Vincent Guidard, Philippe Chambon, Jean-François Mahfouf, Pierre Brousseau, Patrick Moll, Thomas August, Tim Hultberg

Preliminary Assessment of MetOp-Based Temperature and Humidity Statistical Retrievals within the 3D-Var AROME-France Prediction System

BRUNA BARBOSA SILVEIRA,[a] NADIA FOURRIÉ,[a] VINCENT GUIDARD,[a] PHILIPPE CHAMBON,[a] JEAN-FRANÇOIS MAHFOUF,[a] PIERRE BROUSSEAU,[a] PATRICK MOLL,[a] THOMAS AUGUST,[b] AND TIM HULTBERG[b]

[a] *CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France*
[b] *EUMETSAT, Darmstadt, Germany*

(Manuscript received 11 June 2020, in final form 4 January 2022)

ABSTRACT: The main objective of the study is to evaluate the feasibility and benefits of assimilating satellite temperature and humidity soundings (aka Level 2 or L2 profiles), instead of radiances, from the EUMETSAT Advanced Retransmission Service (EARS) into the AROME-France data assimilation system. The satellite soundings are operational forecast-independent retrievals that used the infrared sounder IASI in synergy with its companion microwave instruments AMSU-A and MHS on board the MetOp platforms. In this assimilation study, L2 profiles are used as pseudoradiosondes, discarding vertical error correlations and the instrument vertical sensitivity in the observation operator due to the lack of available averaging kernels. Three assimilation experiments were performed, the baseline (including all satellite radiances except those from IASI, AMSU-A, and MHS sounders), the control (with observations from the baseline plus IASI, AMSU-A, and MHS radiances), and the L2 experiment (with observations from the baseline and L2 temperature and humidity profiles). The assimilation runs cover the periods of the winter 2017 and summer 2018. The forecast skills of the three experiments are gauged against independent analyses and observations. Despite that the vertical observation operator is not accounted for in this study, it is found that L2 profile assimilation does not have a negative impact on 1-h temperature and humidity forecasts, especially in the midtroposphere. Their impacts are comparable in magnitude to the radiance ones in the operational AROME framework, except in terms of temperature and wind fields during winter where the impact is more negative than positive. These findings encourage further investigations.

KEYWORDS: Atmosphere; Remote sensing; Satellite observations; Numerical weather prediction/forecasting; Data assimilation

1. Introduction

Satellite observations are the main sources of observations used in the initialization of numerical weather prediction (NWP) models. These data are assimilated into NWP models together with other observations such as surface synoptic data, radiosondes, aircraft, and radar. Among these satellite data, infrared and microwave sounder data represent the vast majority of the information ingested in global models. In limited-area (regional) models, however, observations from polar orbiting satellites (in the present study only *MetOp-A* and *-B*) can represent only around 5% of the total amount of data (presented later in the text). In this case, the more numerous observations are those of synoptic (34%), radar (33%), and aircraft (17%).

The assimilation of L2 profiles is nothing new. In the early years of atmospheric sounding from space, NWP models were assimilating geophysical quantities (e.g., atmospheric temperature and humidity) retrieved from satellite observations [the so-called Level 2 profiles (L2)]. This was the usual approach, but success was limited, because it led to an influx of information that did not originate from the satellite data, i.e., the errors of the retrieval a priori. In the 1990s (Eyre et al. 2019), the direct assimilation of Level 1 (L1) radiances was generalized as opposed to that of the L2. This is still the paradigm nowadays and remains in the current plans for operational NWP. To understand the failure of the early attempts of L2 assimilation, we recall that the temperature and humidity information, which can be extracted from the satellite data, is of limited vertical resolution. This means that high-vertical-resolution representations of L2 profiles cannot be uniquely determined by the satellite data since there are vertical structures, which are not measured by the satellite instruments. In some L2 retrievals, the loadings along such directions are chosen to be zero, effectively selecting the smoothest profiles consistent with the measurements. In others, these loadings are determined by some auxiliary data (often referred to as "a priori"). In any case, it is harmful if the assimilation is allowed to influence these unmeasured vertical (finer) structures of the model profiles. As mentioned, this can be avoided not only by assimilating L1 instead of L2, but also by using the averaging kernels of the L2 profiles as observation operator in L2 assimilation (or similar approaches to avoid affecting the model along the directions of the retrieval null space).

For quite a while, the assimilation of hyper-spectral radiances (Hilton et al. 2012; Guidard et al. 2011) has had a very positive impact, even though only radiances from a fraction of available channels are assimilated. Recent and future missions constitute computational and scientific challenges for these applications, however, which can still only exploit a subset of the information. On one hand, the radiative transfer model, which is the observation operator in the radiance assimilation, is a major contributor to the overall computational budget in

Corresponding author: Bruna Barbosa Silveira, brunabs.silveira@gmail.com

DOI: 10.1175/MWR-D-20-0189.1

an assimilation run. On the other hand, systematic and random errors in the model background, in the radiative transfer model itself and the limitation of unidentified clouds in the observations, can significantly affect the ability to fit observed radiances with those which are simulated. As a result, measurements have usually less weight in the assimilation or they are discarded over land for surface-sensitive or low-peaking channels, especially those sensitive to humidity. Radiances affected by cloud effects are also removed (over land and sea). It is noteworthy, that the data volume from hyper-spectral sounders will keep increasing in the years to come, with additional and data-denser missions in space, e.g., with twice as many channels for Infrared Atmospheric Sounding Interferometer–Next Generation (IASI-NG) or 100 times more daily observations from Meteosat Third Generation–Infrared Sounder (MTG-IRS), when compared to IASI.

In contrast, as no radiative transfer model is needed, the assimilation of geophysical products can be performed at a lower computational cost. Also, state of the art L2 algorithms exploit the measurements from a large fraction of channels and are also improved in some cloudy situations. L2 profiles (temperature and humidity profiles) have hence the potential to more exhaustively convey information from satellite observations into the system. This could be especially relevant in rapid-cycle convective models, when computational resources, model resolution (horizontal and vertical) and the overall amount of assimilated data may be competing against particularly strong requirements in terms of timing. Following the recent progress on L2 performances, there is a renewed scientific interest in this field (Otkin et al. 2011; Hartung et al. 2011; Migliorini 2012; Coniglio et al. 2019; Hu et al. 2019).

In this context, this study explores the practical requirements and potential benefits of assimilating operational temperature and humidity retrieval profiles from the EUMETSAT EARS-IASI L2 regional service without considering their associated averaging kernels (since they are not disseminated with the L2 profiles). The inclusion of averaging kernels in the disseminated L2 products would imply a major increase of the data volume and EUMETSAT is currently looking into ways of mitigating this problem. These efforts are based both on a lower dimensional representation of the profiles, as PC scores, and the possibility of constructing the averaging kernels at the user side from a few disseminated indices and a static lookup table disseminated only once. In this case, since the vertical sampling of the assimilated profiles is higher than the actual resolution, which can be determined by the retrievals, spurious high-frequency patterns of the retrieval prior can affect the analysis. The former version of the AROME-France model was able to assimilate 124 of IASI channels (Seity et al. 2011). AROME-France model top has been lowered, however, from 1 to about 10 hPa in order to focus on the troposphere as AROME-France is not actually used in the stratosphere. This has thus saved numerous vertical levels, which would be needed to correctly simulate strong stratospheric winds at high horizontal resolution (Brousseau et al. 2016). Channels with a contribution therefore above 10 hPa are not assimilated as their simulations are of a lesser quality given the reduced number of vertical levels in the stratosphere.

The goal of this paper is to evaluate the feasibility and the potential benefits of assimilating L2 profiles (temperature and humidity profiles) in a regional model instead of IASI, AMSU-A, and MHS radiances. This evaluation contributes to the probing of new applications for current operational products from EUMETSAT Polar System (EPS), and to the enhanced understanding of their utilization bearing in mind future hyper-spectral missions of the Infrared Atmospheric Sounding Interferometer–New Generation (IASI-NG) and Meteosat Third Generation–InfraRed Sounder (MTG-IRS).

To achieve the objective of this study, the L2 temperature and humidity profiles were first of all assessed against AROME-France 1-h forecasts, which represent the current state of the AROME-France assimilation system (first-guess). Second, the configuration of the assimilation experiment that uses the L2 profiles instead of radiances (Level 1) in the AROME-France system was defined. Finally a set of 3 experiments was compared in order to evaluate the assimilation of L2 profiles: the baseline, which assimilates the current set of observations in the AROME-France operational version in 2017 and 2018 excepting radiances from IASI, AMSU-A, and MHS sounders, the control experiment (similar to the baseline but also assimilating radiances from IASI, AMSU-A, and MHS sounders) and the L2 experiment (similar to the baseline, but also assimilating L2 profiles). In addition to the results of the assimilation experiments, findings and possible recommendations related to the IASI L2 profiles themselves are given in this paper.

The paper is divided as follows: the second section describes the AROME-France model and L2 profiles, the third section presents the evaluation of the L2 profiles compared to the AROME-France model, which will help to build up the following data assimilation experiments. The fourth section outlines the definition of the experimental set-up and the main results. Finally, the last section provides the conclusions and studies to follow.

2. Model and data

a. AROME-France and data assimilation

Figure 1 presents the domain and orography of the convective-scale AROME-France model. Since 2008, this model is operational at Météo-France (Seity et al. 2011). The AROME-France domain is centered at 47.5°N, 2°E over France and contains 1440 × 1536 points on a Lambert projection (Brousseau et al. 2016). The horizontal resolution of the current AROME-France version is 1.3 km. The levels range from 5 m above the surface to 10 hPa (90 levels). The AROME-France version used in this study was the operational version in 2017 and 2018. A 3D-Var assimilation scheme with hourly cycling provided the initial conditions for the AROME-France model, with a 1-h assimilation time window centered on the round hour (Brousseau et al. 2016), which means that innovations are all assumed as valid at the central time.

The AROME-France system is capable of assimilating different observation types, such as surface stations, buoys, radar

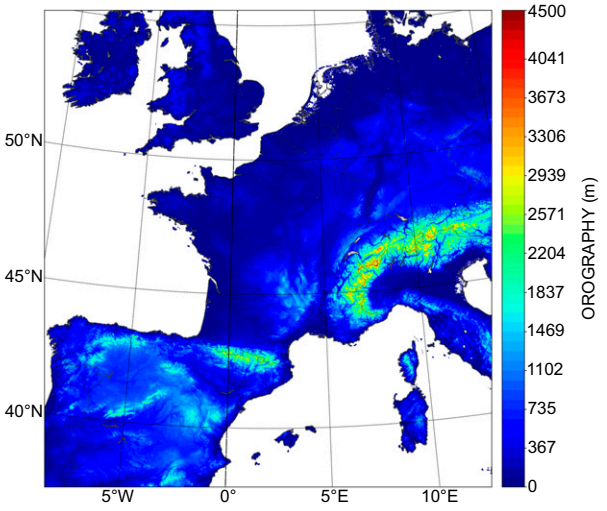

FIG. 1. Coverage of the AROME-France and orography in shaded colors.

measurements (Doppler wind and reflectivity), aircraft, wind profilers, ship, satellite observations, and radiosondes. Radiosondes relative humidity profiles are not assimilated above 300 hPa. Radar reflectivities are assimilated using a 1D method (Wattrelot et al. 2014), which provides relative humidity profiles for the 3D-Var. Satellite observations contain data from microwave and infrared sensors. These sensors are on board polar and geostationary satellites. In AROME-France, the operationally assimilated sensors were at that time MHS on board *NOAA-18* and *-19* and *MetOp-A* and *-B*; AMSU-A on board *MetOp-A* and *-B*, *NOAA-15*, *-18*, and *-19* and *Aqua*; ATMS on board *SNPP*; SEVIRI from *Metosat-11*; GMI (GPM); IASI on board *MetOp-A* and *-B*; SSM/IS (*DMSP-17* and *-18*) and scatterometer (*MetOp-A* and *-B*). The GNSS zenithal total delay data from ground-based stations and atmospheric motion vectors from SEVIRI are also assimilated.

Medium-range forecasts are issued from AROME-France analyses of the assimilation cycle at 0000, 0600, 1200, and 1800 UTC. At 0000 and 1200 UTC the forecast range is 48 h and at 0600 and 1800 UTC, it is 42 h. The forecast fields from the French global model, Action de Recherche Petite Echelle Grande Echelle (ARPEGE; Courtier et al. 1991), are used as the boundary conditions in the AROME-France model.

The radiances, like other observations, are bias corrected and the variational bias correction (Auligné et al. 2007) is used in the AROME-France system for satellite observations. The bias correction coefficients applied to radiance observations are calculated with ARPEGE except those applied to SEVIRI radiances, which are computed in the AROME-France. SEVIRI clear sky radiances averaged over boxes are indeed assimilated in ARPEGE and not raw radiances as in AROME-France. Given their availability over the AROME-France domain, the MetOp observations are assimilated from 0800 to 1200 UTC and from 1900 to 2300 UTC. Horizontal thinning is also applied to radiance observations in the AROME-France system since horizontal observation error correlations are neglected too as a way in which to avoid information redundancy.

b. *EARS-IASI L2 profiles*

The profiles are generated through the EUMETSAT Advanced Retransmission Service, EARS-IASI L2 (https://www.eumetsat.int/ears-iasi), whereby MetOp observations are directly broadcast to local receiving stations and centralized to generate L2 profiles at EUMETSAT headquarters. The thermodynamic profiles (temperature and humidity profiles) are then disseminated through EUMETCast and available to users within less than 30 min from sensing.

The period evaluated in this study comprises 8 months (from July 2017 to February 2018). In this study, the L2 profiles considered are the ones locally received in real time at Lannion-France and one of the latest EARS-IASI L2 profile generation (version 6.3; EUMETSAT 2017a). The L2 profiles are available for AROME-France data assimilation from 0800 to 1200 UTC (for *MetOp-A* and *-B*) and from 1900 to 2300 UTC (only for *MetOp-B*).

The temperature and humidity profiles in the L2 profiles are retrieved with a statistical method using IASI measurements in synergy with companion microwave sensors (AMSU-A and MHS) on board MetOp satellites. In this method, the satellite observations are represented in principal component (PC) scores to form the predictor input vector, together with auxiliary information such as surface elevation, satellite zenith angle, and others. The temperature and humidity profiles are also represented in PCs to form the predictands. The predictants come from ERA-5. This will imply that any model bias will be inherited by the retrieval. In a future study, an online bias correction through VarBC could, however, be considered to prevent this issue. The supervised machine learning was performed using real EPS radiance observations collocated with ERA-5 fields during 2015 and 2016. In total, more than 100 million pairs were used in the training.

The observation space is first divided into classes by application of *k*-means clustering on the IASI and microwave observations. The classification (*k*-means clustering) is based on the measurements and divides the fields of view into classes with similar atmospheric situations. Pixels with large amount of water vapor do not belong to the same classes as pixels with lower amount of water vapor. This classification into relatively homogeneous classes makes the independent linear regressions within each class work well. This clustering into relatively homogeneous classes allows a simple mathematical modeling of the relationship between radiances and geophysical parameters, using a linear regression. The visual inspection at global scales of the classification shows that the air mass observations are clustered in a spatially coherent manner. Extensive routine comparisons to radiosondes confirmed that the retrieval precision reached with this approach is very competitive, with temperature precision typically better than 1 K in the free troposphere and low-level humidity precision within 1–1.5 g kg^{-1} in clear sky (Feltz et al. 2014; Roman et al. 2016; EUMETSAT 2017a). The statistical

relationship is then modeled by simple linear regression in each observation class, achieving overall nonlinear inference between observations and geophysical state through piecewise linear regressions.

In practice, several retrieval instances are invoked and averaged, effectively forming an ensemble statistical retrieval in every pixel. The different instances of the retrieval coefficients are obtained by varying the selection of the measurement PC scores used for the classification. A large part of the individual retrieval errors are independent so that the overall retrieval errors are reduced by the averaging. Measurements in adjacent pixels in an effective field of view (EFOV) are also exploited in the input vector so as to take advantage of geophysical horizontal correlations. Four sets of profiles, corresponding to the four IASI FOVs of an EFOV are retrieved from measurements from four IASI FOVs, nine MHS FOVs, and one AMSU FOV. The spatial variation within the EFOV is determined by the IASI and MHS measurements only. For overcast situations where most of the low-level temperature information comes from AMSU, the retrieval might not be able to fully capture the spatial variation. The method is referred to as Piece-Wise Linear Regression-cube ($PWLR^3$) and is presented and evaluated more in detail in EUMETSAT (2017b). All IASI channels are used in the $PWLR^3$. The retrieval precision, assessed by comparison with radiosondes, shows that the piece-wise linear function implemented by the $PWLR^3$ could be a valuable approximation of the relationship between observations and atmospheric profiles.

The temperature (in K) and humidity (mixing ratio) profiles are sampled on the vertical levels from the surface to 10 hPa and provided along with quality indicators (QIs). The QIs are uncertainty estimates as part of the main regression, also output from $PWLR^3$, related to the precision of the retrieved temperature (in K) and humidity (in dewpoint temperature in K) in the lower troposphere. In addition to the temperature and humidity profiles, a cloud signal is also retrieved from $PWLR^3$. It is referred to as OmC (short for OBS minus CALC), which gives the difference in brightness temperature between observations and clear-sky simulations in atmospheric window channels. It provides an indication of a possible cloud contamination within a pixel. The cloud QIs are equal to the OmC value. The simulations are performed offline on the training set, the retrievals can be trained with OmC values and no forward model simulations are needed for the online application of the retrieval.

The present work used profiles of pressure, water vapor mixing ratio converted in specific humidity using equation $q = r/(1 + r)$ as it is the variable for humidity in AROME-France, the temperature for the evaluation of L2 profiles compared with 1-h forecasts and also for the assimilation of L2 profiles. The mean elevation within a given pixel was used due to differences between AROME-France model and L2 profiles orography values. Profiles having an absolute orography difference larger than 25 m were discarded. The QI for humidity, temperature and cloud were used to select the most accurate data for the L2 profiles evaluation against 1-h forecast and also for assimilation. The evaluation has been undertaken on profiles with QI values less than 2 K for temperature and less than 3 K for dewpoint temperature (i.e., humidity). These QI values were selected after an evaluation of all L2 temperature and humidity profiles against the profiles satisfying these criteria. The comparison of these two datasets showed that there is no loss in profile quality (results not shown). The cloud QI is only used for assimilation when there is more than one profile with in a grid box [thinning process is explained in section 4a(2)] with temperature and humidity profiles passing the above QI criteria.

The L2 profiles are assimilated without accounting for the instrument vertical sensitivity, i.e., without application of the averaging kernels (Rodgers 2000) as observation operator. This study constitutes a first attempt to assimilate, in a simple framework, the temperature and humidity profiles retrieved from satellite observation.

3. Comparison of MetOp-combined retrieval L2 profiles and AROME-France model

The first step of any observation prior to assimilation is comparison against a model counterpart. This assessment helps to identify the characteristics of the main differences between both datasets, thereby providing useful information for the assimilation of L2 profiles. Figure 2 shows the mean of monthly mean and standard deviation values of the differences between L2 profiles and AROME-France 1-h forecasts (i.e., first-guess departure, known hereafter as OMF) over the AROME-France domain for a 8-month period (July 2017–February 2018). First guess is the observation equivalent simulated from the 1-h forecast. The evaluated variables are temperature (Fig. 2a) and specific humidity (Fig. 2b). The temperature mean of monthly means of OMF varies greatly close to the surface and lies between 200 and 300 hPa. Near the surface, these variations range between −3 and −0.5 K (probably related to a difference in the surface characterization). In the upper atmosphere, they range from −0.9 to 0.4 K and can be related to differences in the tropopause position. Between 800 and 600 hPa the mean of monthly means of OMF is negative and close to zero (Fig. 2a). The mean of monthly standard deviation of OMF is around 1 K above 800 hPa and up to 300 hPa, near the surface, it is 2.3 K. Between 300 and 200 hPa, there is an increase in both the mean of monthly mean and standard deviation values of OMF. The mean of monthly standard deviation values vary from 1 to 1.5 K.

The specific humidity mean of monthly mean OMF are negative for most levels, which means that the background contains more humidity than do the L2 profiles. Near the surface, the mean of monthly mean values of OMF is negative around -0.04 g kg^{-1}. A monthly evaluation shows that the standard deviation of OMF presents a strong variation linked to the absolute atmospheric moisture content seasonal variation (represented by dashed red and green lines in Fig. 2b). This means that moister atmospheres are associated with larger deviation.

The statistics of L2 OMF from two periods (July–September 2017 and January–February 2018) are also compared

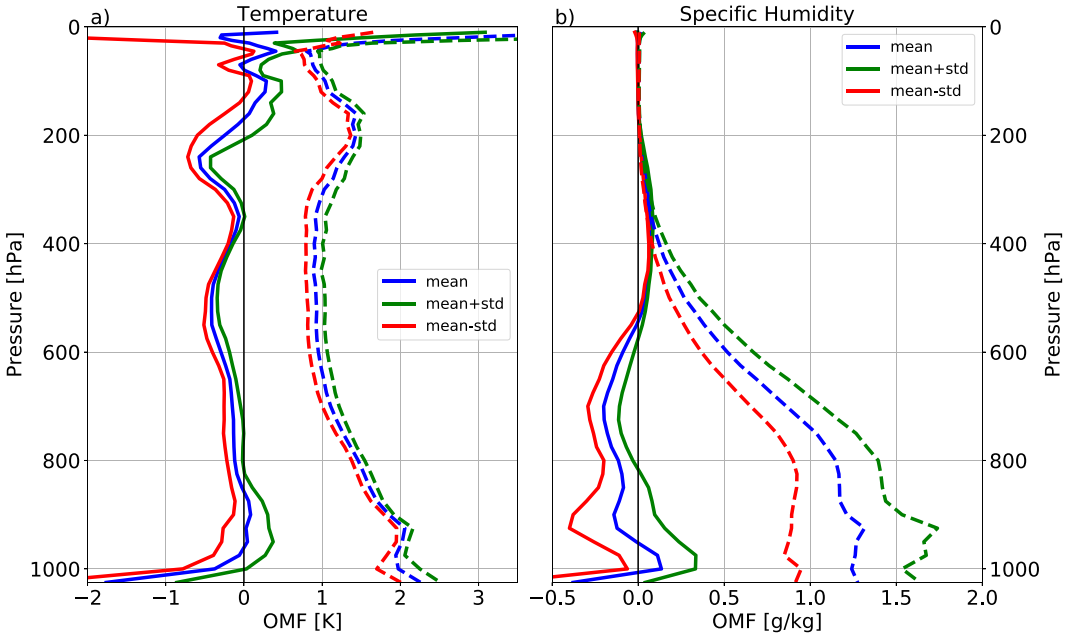

FIG. 2. L2 profiles minus AROME-France first-guess statistics for (a) temperature and (b) specific humidity, calculated over the 8 months (July 2017–February 2018). Mean of monthly means differences are in solid lines and mean of monthly standard deviation of differences are in dashed lines. Blue lines represent mean of monthly difference mean and mean of monthly standard deviation values. Red solid lines depict mean of monthly means plus one standard deviation in the monthly means and green solid lines represent mean of monthly means minus standard deviation in the mean. Red dashed lines depict the mean of monthly standard deviation plus one standard deviation in the monthly standard deviation and green dashed lines represent mean of monthly standard deviation minus the standard deviation in the monthly standard deviation.

against the statistics of radiosondes and aircraft OMF assimilated in the baseline experiment (detailed in the next section). The aircraft data also provide vertical temperature profiles close to airport locations. Figure 3a shows the mean OMF for temperature, for radiosondes and L2 profiles. Close to the surface, OMF are of opposite sign (Fig. 3a), the mean OMF is 0.06 K for radiosondes and −0.35 K for the L2 profiles. The absolute values of the amplitudes of the mean OMF display similar behavior between radiosondes, aircraft and L2 profiles. Above 700 hPa, the standard deviations of OMF of the three observation types become closer. The aircraft standard deviation (dashed blue line) is maximal (2 K) near 925 hPa. Close to the surface, the L2 standard deviation is above 2 K whereas the one for radiosondes reaches only 1.3 K. Both specific humidity mean OMF values are in good agreement over the whole atmosphere (Fig. 3b). The standard deviations reveal differences below 700 hPa.

The L2 profiles OMF shown in these evaluations are close to the L2 profile evaluation compared with radiosondes shown in the validation report (EUMETSAT 2018) and described below. The evaluation made by EUMETSAT selected radiosondes over Europe within 3 h and 50 km from the *MetOp-A* and *-B* pixels and cloud-free pixels. The evaluation period ranges from 23 December 2017 to 13 February 2018 (this period is within the period used here to compute the OMF statistics of L2 profiles). Between 800 and 400 hPa the mean differences in temperature between L2 profiles and radiosondes are similar to the results found for the L2 mean OMF statistics, which are −0.5 to −0.75 K for *MetOp-A* and *MetOp-B*, respectively. Close to the surface there is an opposite signal. The standard deviation of differences for temperature has a similar shape to those observed for the mean OMF of L2 profiles, with a decrease of the differences from 1000 hPa up to around 450 hPa, an increase above this level and a reduction above 150 hPa.

The mean differences between L2 profiles and radiosondes for water vapor are positive and higher close to the surface (around 0.5 g kg^{-1}) for *MetOp-B*, whereas for *MetOp-A* they become negative at 900 hPa and are again positive above this level (Fig. 2). The mean OMF values of L2 profiles for water vapor are negative up to 600 hPa. Above this level, there is a behavior similar to the differences between radiosondes and L2 profiles. The standard deviations of the differences for the L2 water vapor profiles are similar for the two comparisons. The small differences between the evaluation made by EUMETSAT and the one performed in this study can be related by the different period analyzed in this paper, which included winter and summer months.

The evaluation of L2 profiles (shown above) in terms of first-guess departure statistics concerning temperature and humidity showed similar results when compared against the radiosonde and aircraft data. L2 profiles appear therefore as suitable for assimilation in AROME-France with inflated

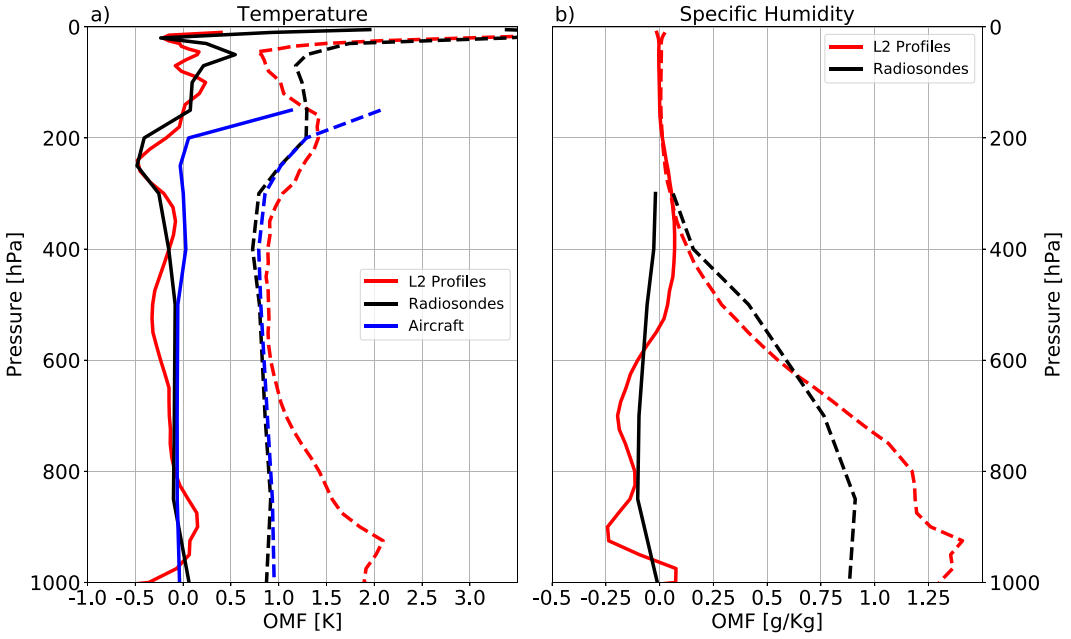

FIG. 3. L2 profiles, radiosonde, and aircraft minus AROME-France first-guess (OMF) statistics for July–September 2017 and for January–February 2018. Mean differences are in solid lines and standard deviations of differences are in dashed lines. Black lines: radiosondes; red: L2 profiles; and blue: aircraft (AIREP). (a) Temperature and (b) specific humidity.

errors (presented in the next section) on specific levels with respect to radiosondes and aircraft data.

4. Results of the assimilation experiments

a. Experimental setup

The L2 profiles were assimilated in the AROME-France 3D-Var system during two separate 2-month periods, one during the summer from 15 July to 15 September 2017, the other during winter from 1 January to 28 February 2018. Three experiments, summarized in Table 1, were performed during each period. In the baseline experiment, all observations considered in the operational configuration, except IASI, AMSU-A and MHS radiances (L1 product) were assimilated. In the control experiment, the same observations as in the baseline experiment plus IASI radiances from Lannion–France (locally acquired), AMSU-A and MHS radiances from EUMETSAT but no evening *MetOp-A* radiances were assimilated. These data are not received in Lannion due to the Atlantic anomaly (the direct broadcast of *MetOp-A* is switched off in the evening overpasses). The AMSU-A and MHS observations from only *MetOp-A* and *-B* satellites were assimilated in the control experiments. The AMSU-A and MHS observations from the other satellites were removed from the assimilation. In the L2 experiment, baseline observations plus the L2 profiles (temperature and specific humidity profiles) from Lannion were assimilated, excepting L2 profiles from evening *MetOp-A*. As a consequence, the MetOp observations are assimilated from 0800 to 1200 UTC and from 1900 to 2300 UTC (only *MetOp-B*).

TABLE 1. Experiment configurations.

Expt	Configuration
Baseline	No IASI, AMSU-A, and MHS data
Control	Baseline + IASI from Lannion (only), AMSU-A, and MHS data from EUMETSAT; no *MetOp-A* in evening
L2	Baseline + L2 profiles from Lannion; *MetOp-A* and *-B* in the morning and only *MetOp-B* in the evening

1) L1 ASSIMILATION DETAILS

IASI, MHS and AMSU-A brightness temperatures, together, represent between 3% and 4.7% of the data assimilated into the AROME-France system over winter and summer periods, respectively. These percentages are a function of weather conditions; cloudy or clear sky. The observation operator used by AROME-France system to simulate satellite radiances is based on the RTTOV model (Saunders et al. 2018). The horizontal thinning applied to AMSU-A is 100 km and to IASI and MHS observations is 80 km. Figure 4 shows the observation position (of at least one assimilated channel) of IASI (black squares), AMSU-A (red diamonds), and MHS (yellow circles) assimilated in the control experiment within a 1-h assimilation time window (at 1000 UTC 1 January 2018). IASI, AMSU-A and MHS present a good horizontal distribution and are not necessarily collocated. Figure 4 also shows cloud-top pressure estimated from SEVIRI observations (shaded areas). It appears that IASI observations can be

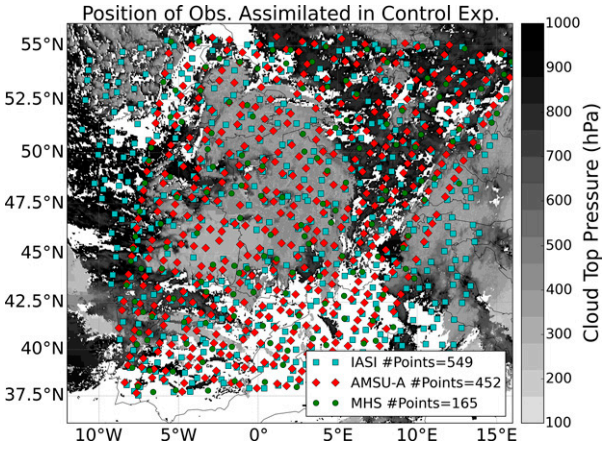

FIG. 4. Horizontal distribution of assimilated L1 observations at 1000 UTC 1 Jan 2018 from IASI (cyan squares), AMSU-A (red diamonds), and MHS (green disks). Cloud-top pressure (hPa) estimated from SEVIRI observations in shaded colors for the same date.

assimilated in some cloudy situations (depending on cloud type). There are 47.5% of the assimilated IASI pixels in this time window (between 0930 and 1030 UTC 1 January 2018) with a cloud fraction higher than 90% (AVHRR information) and only 27% of the IASI pixels are associated with a cloud fraction between 0% and 9%.

In this AROME-France configuration, over land, only 8 (water vapor) IASI channels peaking between 600 and 200 hPa are assimilated. Over sea, 44 IASI channels are assimilated (20 water vapor, 20 temperature, and 4 surface channels). In comparison, ARPEGE model can assimilate 129 IASI channels. Jacobians of the IASI channels assimilated in this AROME-France system over land (8 channels) and over sea (44 channels) are presented in Figs. 5a and 5b, respectively. Four AMSU-A channels (5, 6, 7, and 8) and three MHS channels (3, 4, and 5) are assimilated into AROME-France model (over land and sea). Over high orography channels 5 (>500 m) and 6 (>1500 m) from AMSU-A and all channels from MHS (channel 4 and 5 orography > 1000 m and channel 3 orography > 1500 m) are not assimilated, however. Jacobians of these channels are shown in Figs. 5c and 5d, respectively, for a given atmospheric profile. Over sea, in the low troposphere, the information assimilated in the AROME-France model is derived essentially from satellite data.

2) L2 ASSIMILATION DETAILS

The L2 profiles are assimilated as pseudoradiosondes in the AROME-France system. Only the diagonal of observation-error covariance matrix is considered for the L2 profiles. The same consideration is made for the radiosondes/aircraft. Prior to being assimilated the L2 profiles observation error (σ_{oL2}) has to be specified. The L2 profiles QIs cannot be used as L2 profiles observation error (σ_{oL2}) because there is only one QI value per profile, but the assimilation scheme requires one value of error per observation level (L2 profile levels). The L2 observation errors (σ_{oL2}) were determined from radiosondes and aircraft characteristics. The AROME-France operational version used in this study assumes that radiosonde and aircraft have the same observation error (σ_{oR}) for temperature. As mentioned in the previous section, the evaluation between L2, radiosondes and aircraft data provide useful information for the estimation of L2 profiles observation error (σ_{oL2}).

The L2 profiles mean and standard deviation of first-guess departure (OMF) are larger than radiosondes and aircraft values on several levels. As a preliminary choice, the estimated error of L2 temperature profiles is set to 1.2 times that of radiosondes. In the AROME-France system, observation errors for humidity (radiosonde) are specified in percent of relative humidity. The radiosonde humidity observations are not assimilated above 300 hPa but their observation errors are specified. In this way, the observation error (σ_{oL2}) values for the L2 profiles are empirically set to 15% of the relative humidity. In comparison, the error specified for the humidity from radiosonde is 12% of the relative humidity. The L2 profiles (red lines) and radiosondes (black lines) observation errors (σ_o) are presented in Fig. 6a for temperature and Fig. 6b for specific humidity profiles. Figure 6 also presents the observation error diagnosed for radiosondes (black dashed lines) and L2 profiles (red dashed lines) using the method described in Desroziers et al. (2005). This diagnostic was performed from a 1-month (January 2018) assimilation experiment to evaluate the prescribed observation error (σ_{oL2}) choice. The diagnostics show that the observation error estimated for temperature and specific humidity profiles from radiosondes are overestimated as the errors diagnosed from Desroziers et al. (2005) are significantly smaller. The same feature is observed for the L2 temperature and specific humidity profiles; however, the ratio between the radiosonde and the L2 diagnosed errors (dashed lines) is higher than that between the radiosonde (black continuous line) and L2 (red continuous line) prescribed errors for temperature. This evaluation shows that the prescribed L2 profile observation errors (σ_{oL2}) values are reasonable for a first approach.

A thinning procedure is also necessary for L2 data to avoid representativeness errors, which are not currently taken into account in the AROME-France 3D-Var. The inversion process introduces a correlation between the observation errors of the retrieved profile, which justified large thinning for the L2 profiles. After a trial-and-error process, the value for the L2 observations was set to 160 km. The L2 profile having the best QI for temperature, humidity and cloud is selected within a 160 km × 160 km box. The thinning algorithm in AROME-France presents a known limitation concerning the minimal distance between two L2 profiles, which can be smaller than 160 km when selecting two different boxes (but it is also the case for all other observations). Figure 7 exhibits the horizontal distribution of L2 profiles (black squares) assimilated at 1000 UTC 1 January 2018 where at least one level datum per profile is assimilated (black squares). A drawback of the current thinning can be observed at various locations, for example over northern Italy, where two observations are closer than the 160 km distance chosen in the algorithm. Figure 7 also shows cloud-top pressure

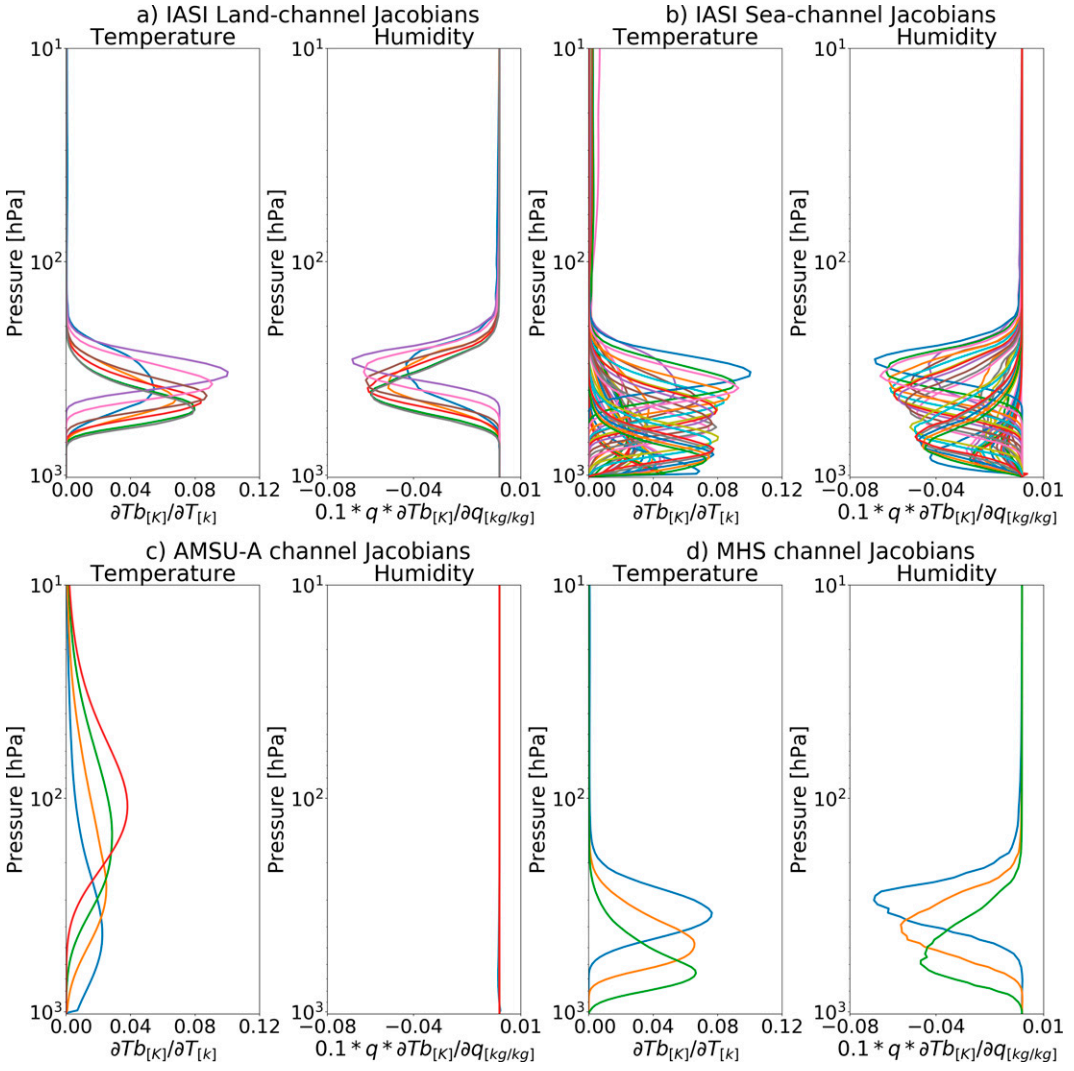

FIG. 5. IASI channels Jacobians (a) over land and (b) over sea. (c) AMSU-A channels 5, 6, 7, and 8 Jacobians and (d) MHS Jacobians from channels 3, 4, and 5.

estimated from SEVIRI observations (shaded areas). It appears that the L2 profiles are assimilated preferentially in clear sky and low cloud areas. There are 79.4% of the profiles with a cloud QI less than 1 K, corresponding to the best cloud QI value.

L2 profiles consist of 109 levels below 10 hPa (corresponding to the AROME-France model top). However, the L2 vertical levels are not equally spaced in pressure and the levels are not the same for all profiles. A first vertical thinning was applied keeping one level in three with a maximum of 36 levels, even if it is not the optimal way to make the level selection and other ways of selection, as the using the averaging kernels, could be applied in future studies. One observation is assimilated per level when the corresponding OMF value is less than a defined threshold (the same as that used for radiosondes).

The evaluation of L2 profiles has been performed separately over sea, over land and at high altitudes (results not shown here). In the following study, we present statistics computed across all surfaces. The L2 experiment makes use of three filters based on OMF statistics over the different surfaces, which also avoid some discrepancies observed in the L2 profiles evaluation. These filters are listed in Table 2. Due to an imperfect knowledge of surface properties, the lowest part of the profile is discarded from the assimilation. More precisely, when the orography is below 1 km, data above 900 hPa level are assimilated, over 1 km, then, data between 700 and 10 hPa are assimilated.

It is important to mention that it is very hard to find settings to make a fair comparison between the control and the L2 experiment (where there are many degrees of freedoms, such as different vertical and horizontal thinning, observation errors, quality control). Each configuration of the control and L2 experiments may be optimized therefore by either improving the radiance assimilation or the L2 assimilation.

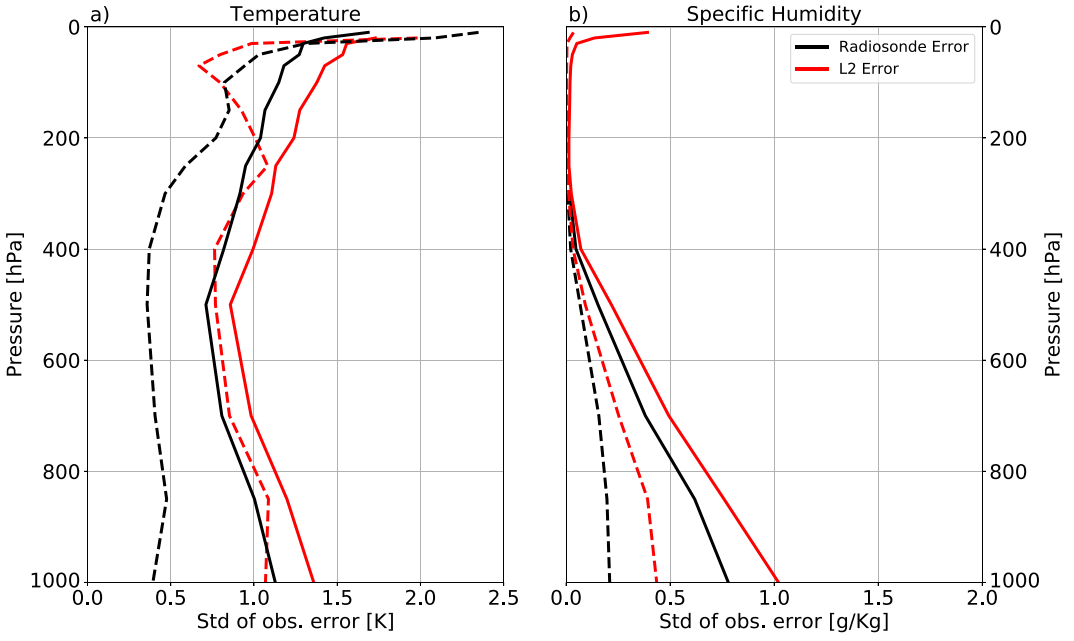

FIG. 6. Specified L2 profiles and radiosondes observation errors (continuous lines) for (a) temperature and (b) specific humidity for January 2018. Dashed lines represent the diagnosed errors using the method described in Desroziers et al. (2005).

Evaluations were carried out in order to investigate the impact of assimilating L2 data instead of L1 on the analyses and forecasts of the AROME-France model. The assessment is presented in two parts: the first one concerns the impact of the use of L2 profiles on the assimilation of other observations. This was evaluated in terms of OMF statistics (mean and standard deviation), which provide the quality of the 1-h forecast against available observations used to generate the analyses. It means that the closer to zero the OMF values are, the better the quality of the forecast. The second type of evaluation examines longer range forecast skill scores in order to assess where (levels and forecast ranges) the assimilation of L2 profiles and L1 products have an impact.

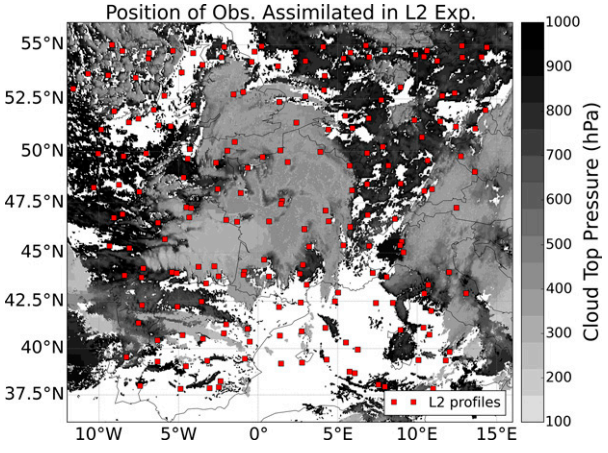

FIG. 7. Spatial distribution of L2 profiles (red squares) at 1000 UTC 1 Jan 2018. Cloud-top pressure (hPa) estimated from SEVIRI observations in shaded colors for the same date.

b. *Impact on the statistics of the other assimilated observations*

The first evaluation concerns the number of observations assimilated in each experiment and for each period. Table 3 shows the number of observations assimilated for each observation type. Excepting radar and radiosonde data, the summer and winter experiments assimilate a similar number of observations (excluding L2 and L1 data from IASI, AMSU-A, and MHS sensors). When both periods are considered, the control experiment assimilates more radiosonde data and radar data when compared with the baseline experiment. However, the relative differences between the control and baseline experiments do not exceed 0.7‰ and 3.5‰ (seventh column in Table 3) for radiosonde and radar data, respectively. When the L2 experiment is compared with the baseline, a larger difference is observed in the number of radar and radiosonde observations assimilated, but the difference between the L2 and baseline experiments does not exceed −3.2‰ and 2.9‰, respectively. The total number of assimilated channels from IASI, AMSU-A, and MHS represent 3.88% of all assimilated data in the control experiment

TABLE 2. Filters applied to L2 profiles.

Region	Filter
Sea	Use data only above level 1000 hPa
Land, orography below 1 km	Use data only above level 900 hPa
Land, orography above 1 km	Use data only above level 700 hPa

TABLE 3. Data assimilated per observation type (ObsType).

ObsType	Baseline	Control	L2	(Control − baseline)	(L2 − baseline)	Diffcontrol (‰)	DiffL2 (‰)
Summer (59 days)							
Synoptic	10 915 891	10 915 809	10 923 776	−82	7885	0.00	0.72
Aircraft	5 828 096	5 827 270	5 827 505	−826	−591	−0.14	−0.10
Satellite	84 344	84 280	84 292	−64	−52	−0.75	−0.61
Drift buoys	61 464	61 464	61 500	0	36	0.0	0.58
Radiosondes	5 780 195	5 780 705	5 796 501	510	16 306	0.09	2.82
Profiler	936 108	936 070	936 542	−38	434	−0.04	0.46
Scatterometers	75 088	75 210	75 210	122	122	1.62	1.62
Radar	7 689 996	7 716 816	7 689 962	26 820	−34	3.49	0.00
GNSS	670 595	670 581	671 187	−14	592	−0.02	0.88
L2 profiles	—	—	2 503 509	—	—	—	—
Radiances[a]	3 901 604	3 900 941	3 903 876	−663	2272	−0.16	0.58
IASI, AMSU-A, and MHS	—	1 768 708	—	—	—	—	—
Winter (63 days)							
Synoptic	10 213 744	10 213 743	10 214 111	−1	367	−9.79 E−5	0.03
Aircraft	4 566 890	4 566 980	4 567 130	90	240	0.02	0.05
Satellite	61 134	61 200	61 188	66	54	1.08	0.88
Drift buoys	48 764	48 768	48 766	4	2	0.08	0.04
Radiosondes	5 329 915	5 333 663	5 329 311	3748	−604	0.70	−0.11
Profiler	700 498	700 470	700 452	−28	−46	−0.03	−0.06
Scatterometers	68 694	68 694	68 694	0	0	0.0	0.0
Radar	12 396 136	12 439 244	12 356 442	43 108	−39 694	3.48	−3.20
GNSS	605 402	605 405	605 406	3	4	0.00	0.00
L2 profiles	—	—	2 275 229	—	—	—	—
Radiances[a]	3 022 002	3 022 499	3 022 310	497	308	0.16	0.10
IASI, AMSU-A, and MHS	—	1 179 720	—	—	—	—	—

[a] ATMS, SSMIS, SEVIRI, GMI, and MWHS2.

(considering both periods). L2 data represents 6.14% of all observations in the L2 experiment for both periods. This result is expected, as more data are found in L2 profiles. In other words, the information comes from 36 levels and 2 variables. The IASI, AMSU-A, and MHS observations represent 51 channels (maximum) over sea and 15 channels (maximum) over land. Even if more L2 profile than radiance observations are assimilated, it is not an indication that they provide additional independent information.

The assimilation of L2 profiles also can have a certain impact on the mean and standard deviation of OMF values of other observation types. Deviations most impacted by the L2 assimilation are the temperature profiles from aircraft and radiosondes, specific humidity profiles from radiosondes and relative humidity derived from radar reflectivities. These observations represent the most numerous ones over land (Table 3) and their OMF can provide evidence about the impact of assimilating the L2 profiles on the background fields (1-h forecast). As differences between both experiments are not great, a significance test (Student's *t* test with 95% of confidence) is necessary to evaluate whether or not these differences are relevant. A significance test (Student's *t* test with 95% of confidence) was applied to the results, the reference experiment being the baseline. Differences exceeding the threshold associated with the 95% level of confidence are assumed to be statistically significant.

Figure 8 shows the vertical profiles of mean OMF values for temperature (aircraft and radiosonde observations). Generally, the mean OMF of temperature from aircraft varies between −0.14 K (winter period) and 0.12 K (summer). Figures 8a and 8b show the mean of OMF for temperature from aircraft during the summer and winter periods, respectively. In the middle atmosphere, it is possible to notice a reduction of the mean OMF values, and this improvement is statistically significant at some levels between 700 and 400 hPa (plain blue triangles). The control experiment also improves the mean OMF of temperature from aircraft mostly between 800 and 600 hPa (plain black triangle). A similar behavior cannot be ascertained above 400 hPa over both periods and for both control and L2 experiments. Below 925 hPa, there is a statistically significant degradation in the L2 experiment. The mean OMF of temperature radiosondes is improved in the L2 and control experiments between 800 and 300 hPa during winter (Fig. 8d). In summer (Fig. 8c), the L2 experiment presents a statistically significant improvement between 500 and 150 hPa, but also a significant degradation (empty blue downward triangle) below 600 hPa.

The normalized standard deviation of OMF from aircraft (Figs. 9a,b) and radiosondes (Figs. 9c,d) are also evaluated. The normalized standard deviations of OMF for temperature for aircraft of L2 (blue lines) and control (black lines) experiments are very similar. The differences between the experiments (control or L2) and the baseline, however, are statistically significant with a degradation in the standard deviation of OMF (downward triangles), which is more

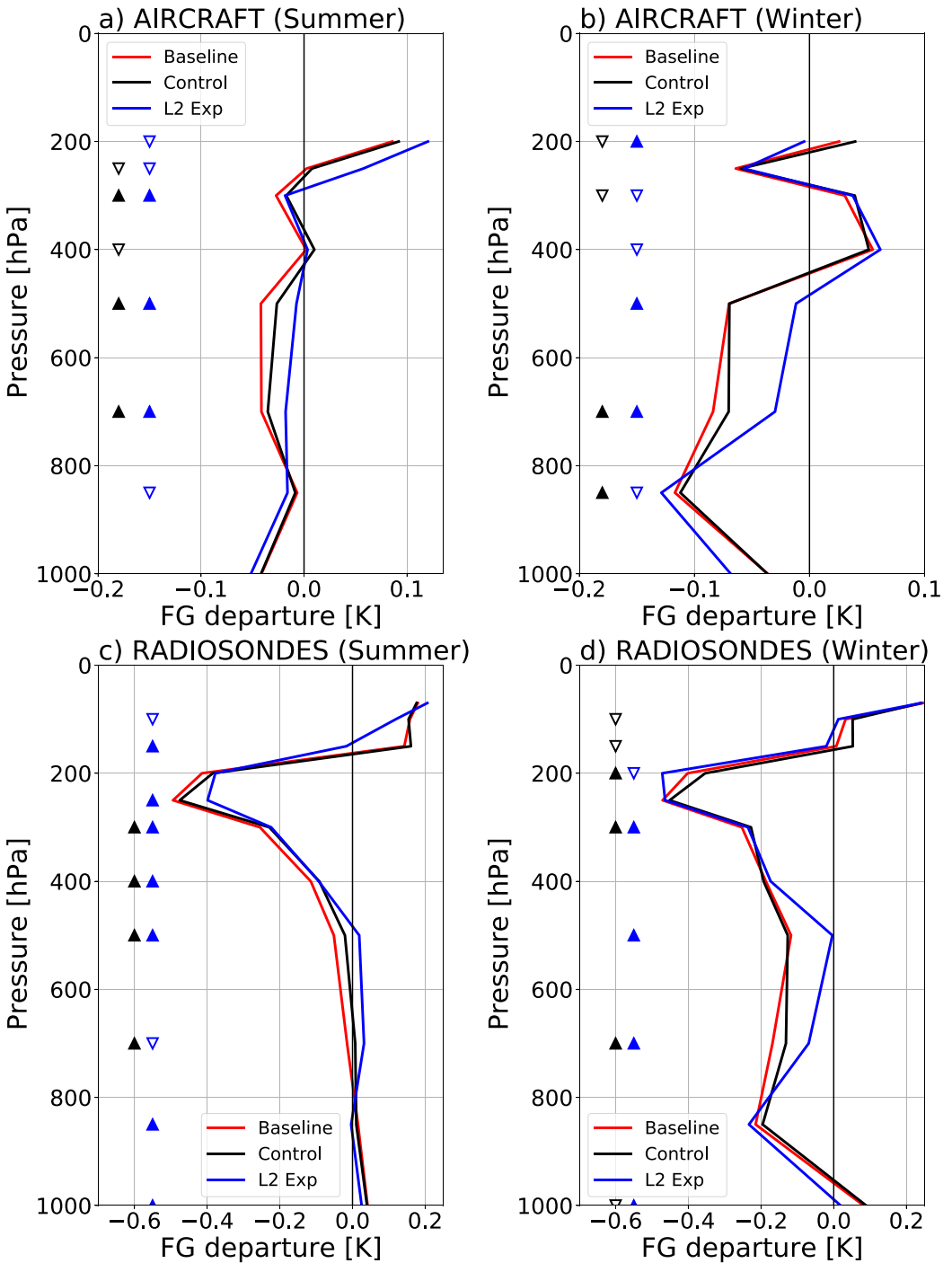

FIG. 8. Mean first-guess departure (OMF) of temperature from aircraft and radiosondes. The blue lines represent L2 experiment, red lines represent the baseline, and the black lines represent the control experiment. Plain black upward-facing triangles mean that the control experiment is better than the baseline with 95% confidence (Student's *t* test). Empty black downward triangles mean that at the level the baseline is better than the control experiment with 95% confidence (Student's *t* test). Blue triangles have the same meaning as the black ones for the L2 experiment. (a) Aircraft over summer period, (b) aircraft over winter period, (c) radiosonde over summer period, and (e) radiosonde over winter period.

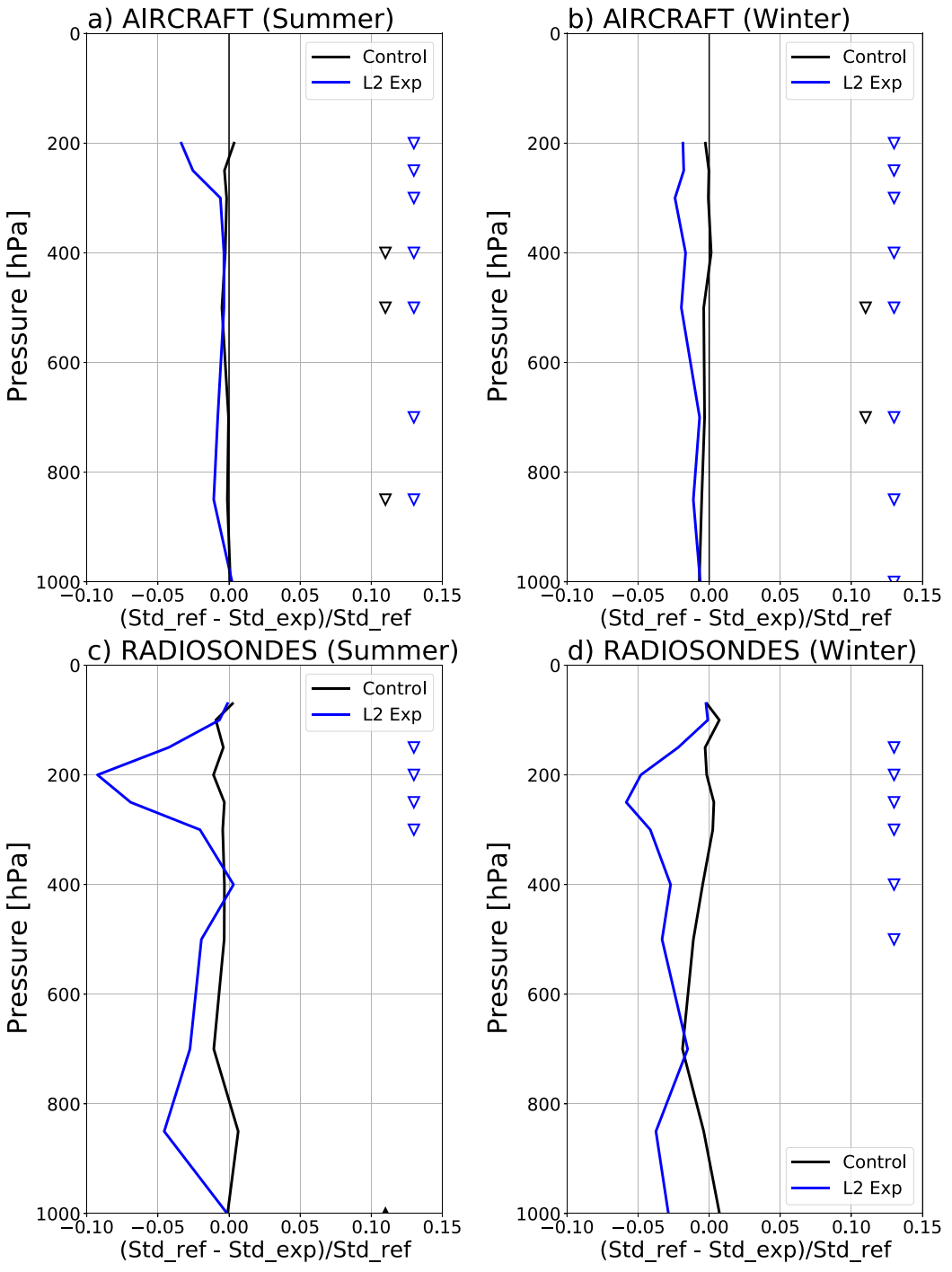

FIG. 9. As in Fig. 8, but for normalized OMF standard deviation where Std_exp is the standard deviation of OMF for the L2 or control experiments and Std_ref is the standard deviation of OMF for the baseline. The blue lines represent L2 experiment and the black lines the control experiment. The triangles have the same meaning as in Fig. 8.

obvious above 400 hPa and between 900 and 600 hPa. The normalized standard deviation of OMF from radiosondes show a degradation in almost all profiles in the L2 experiments where as there is no signal for the control. This degradation is quite obvious above 300 hPa, with statistically significant differences between the L2 experiment and the baseline (empty blue downward triangles in Figs. 9c,d). It is important to mention, however, that the normalized standard deviation of OMF is less than 0.05 in almost every case.

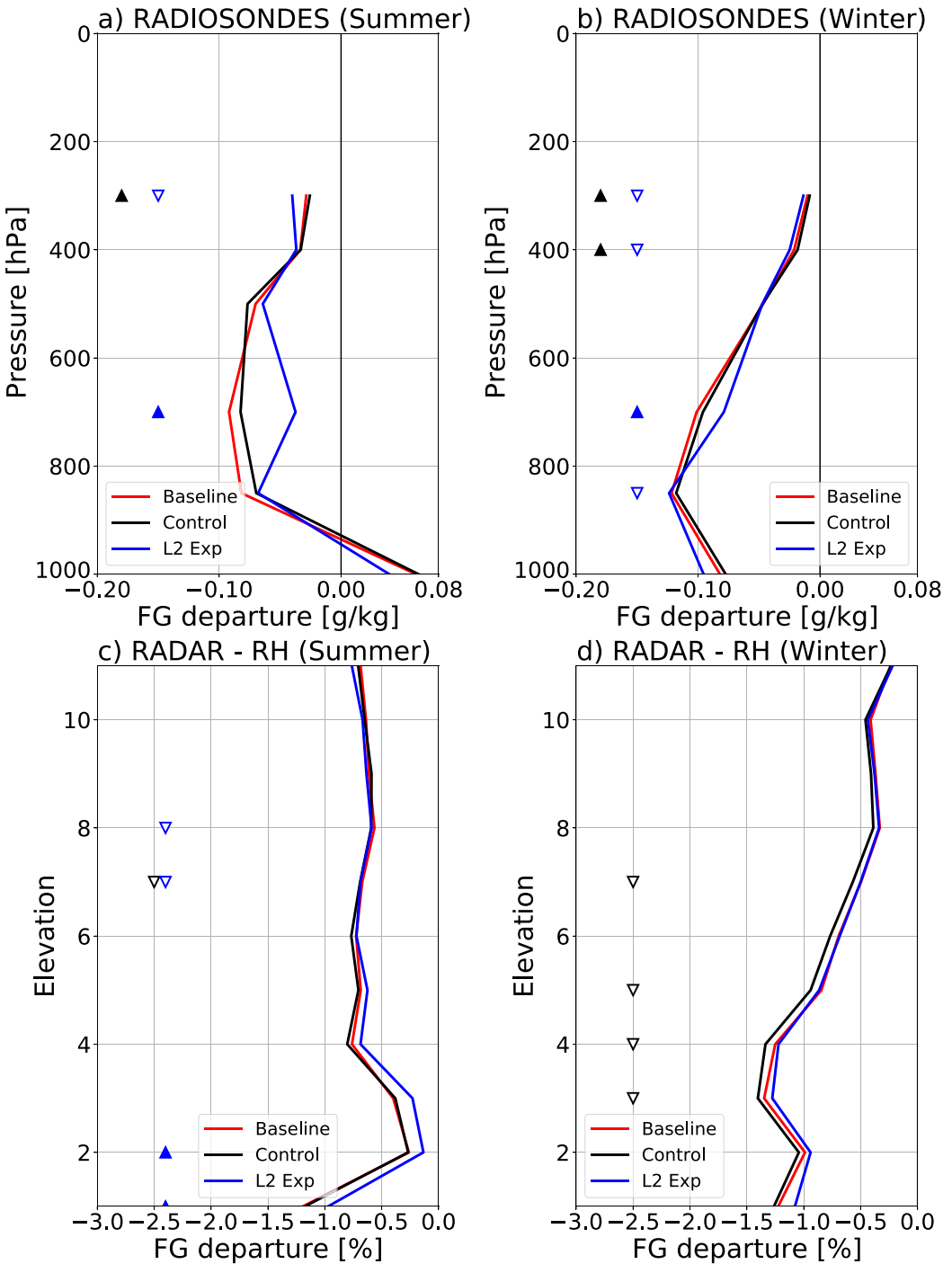

FIG. 10. As in Fig. 8, but for (a),(b) specific humidity from radiosondes and (c),(d) relative humidity (RH) retrieved from radar data.

The mean OMF of radiosonde specific humidity and retrieved relative humidity from radar reflectivity observations are presented in Fig. 10. It is noteworthy that humidity from radiosondes is only assimilated up to 300 hPa. Generally, in the L2 experiment (blue lines), the OMF is reduced in the lower troposphere for both observation types. Above 400 hPa, the values for radiosondes (Figs. 10a,b) present a degradation (improvement) in the L2 (control) experiments with respect to the baseline. These differences are statistically significant at 300 hPa (for both periods) and at 400 hPa (only for winter

period). The L2 experiment improves the humidity first-guess departure of radiosondes between 800 and 600 hPa during the summer and winter periods (Figs. 10a,b), and at 700 hPa this improvement is statistically significant. In this region no water vapor channels are assimilated. Close to the surface, the L2 experiment improves the radiosonde first-guess departure and the control experiment shows an improvement between 850 and 500 hPa during the summer and winter period; however, these differences are not statistically significant.

The vertical information of OMF for radar retrieved relative humidity (Figs. 10c,d) is provided in terms of elevation associated with the scanning mode of each radar. Only during the summer (Fig. 10c) the L2 experiment (blue line) presents an improvement statistically significant up to the elevation 2 and a degradation statistically significant at elevations 7 and 8. The control experiment (black lines) presents a statistically significant degradation in the OMF values at elevation 6 during the summer and up to elevation 7 during the winter period (Fig. 10d).

The normalized standard deviation of OMF for radiosonde specific humidity is displayed in Figs. 11a and 11b. The three experiments present very similar statistics, differences are less than ±0.04 and are not statistically significant. The L2 experiment shows an improvement in terms of OMF standard deviation for radar retrieved relative humidity (Figs. 11c,d) during both periods, and this improvement is statistically significant for almost all elevations (plain blue triangles). An opposite behavior, with differences statistically significant for some elevations, can be seen in the control experiment.

A similar impact study on the assimilation of IASI L2 profiles (retrievals from IASI only) into the European Centre for Medium-Range Weather Forecasts (ECMWF) global data assimilation system focus on data over sea only showed a positive impact in the radiosondes humidity first-guess departures. This was observed mainly when only conventional and AMSU-A were assimilated, Salonen et al. (2019).

The first-guess departure statistics indicate that the L2 profiles can be assimilated into the AROME-France model since the quality of the short-range forecasts is almost unchanged for temperature and humidity fields. The assimilation of the L2 temperature and specific humidity profiles helps to decrease the magnitude of the mean first-guess and analyses departure of other observation types in the midatmosphere (during the assimilation cycle). The observation types evaluated in this analysis were temperature from radiosonde and aircraft, specific humidity profiles from radiosondes and relative humidity derived from radar reflectivities. Figures for analyses departure results are not shown in this paper.

c. *Forecast verification*

1) UPPER-AIR VERIFICATION

For the upper-air verification, analyses from the ECMWF global model, which has a lower resolution (0.5°) than the AROME-France model, are used as reference since they present a good spatial and temporal coverage. This verification was performed for the forecast started at 0000 and 1200 UTC for the summer and winter periods, the 48 h of forecast were evaluated. Only the figures for the forecast initialized at 0000 UTC are shown, however.

The relative differences of temperature root-mean-square error (RMSE) in percent (%) are presented in Figs. 12a and 12g for the forecast initialized at 0000 UTC for summer and winter period for the L2 experiment, respectively. For the summer period, it can be seen that there is a positive (3.5%) relative difference of temperature RMSE at 1000 hPa, which means that the L2 experiment is better than the baseline experiment when compared with the ECMWF's analyses. The L2 profiles are assimilated at 1000 hPa only over sea; however, a L2 profile observation assimilated from 900 hPa can impact variables near the surface through the vertical correlations of background error covariances. These relative differences are statistically significant (horizontal lines in Fig. 12a) up to 12 h. In the wintertime, however, there is a degradation (−3.5%) at 925 and 850 hPa, which is statistically significant. In the 6-h forecast range the relative differences are negative (−3.5%) at 200 hPa for two periods; however, for the winter period they are also observed at 250 and 150 hPa and up to 36 h of forecast (Fig. 12g) these differences are statistically significant. This degradation at 200 and 250 hPa is also observed in the OMF of radiosondes and aircraft (Fig. 8). Concerning temperature, between 700 and 400 hPa, the L2 experiment presents a positive/neutral impact at 6 h forecast, and this impact is also observed in the mean OMF of aircraft and radiosondes. The other levels and forecast ranges do not present a definite behavior. The relative differences of temperature RMSE for the control experiment present a neutral impact in the forecast started at 0000 UTC for summer and winter (Figs. 12d,j, respectively).

Some characteristics observed in the temperature forecast are also observed in the wind intensity scores of the L2 experiment. A degradation can be noted at 200 and 150 hPa in the first 6 h forecast for the summer period (Fig. 12b) and up to 24 h (Fig. 12h) for that of winter, at some forecast ranges and levels the relative differences are statistically significant. The temperature and wind degradation can be explained by the fact that both dynamical fields are linked with different relationships, especially near the tropopause, which imply cross-correlation in background errors for temperature and wind. In addition, during the forecast, temperature and wind forecast errors are strongly linked by geostrophism, thermal wind and baroclinic interaction over the midlatitude. This means that if the temperature field is worse in the analysis and in the subsequent forecast, the wind forecast field will be also impacted and degraded. The wind intensity forecast for the control experiment does not present a well-defined behavior (Figs. 12e for summer and 12l for winter periods), as observed for the temperature forecast.

The relative differences of relative humidity RMSE in % are presented in Fig. 12c (summer) and Fig. 12i (winter) for the L2 experiment. It is possible to notice a positive and statistically significant impact at 100 hPa for summer and at 150 and 100 hPa for winter at all forecast ranges. At 200 hPa during summer and at 250 hPa during the winter there is a statistically significant negative impact up to 12 h of forecast. The

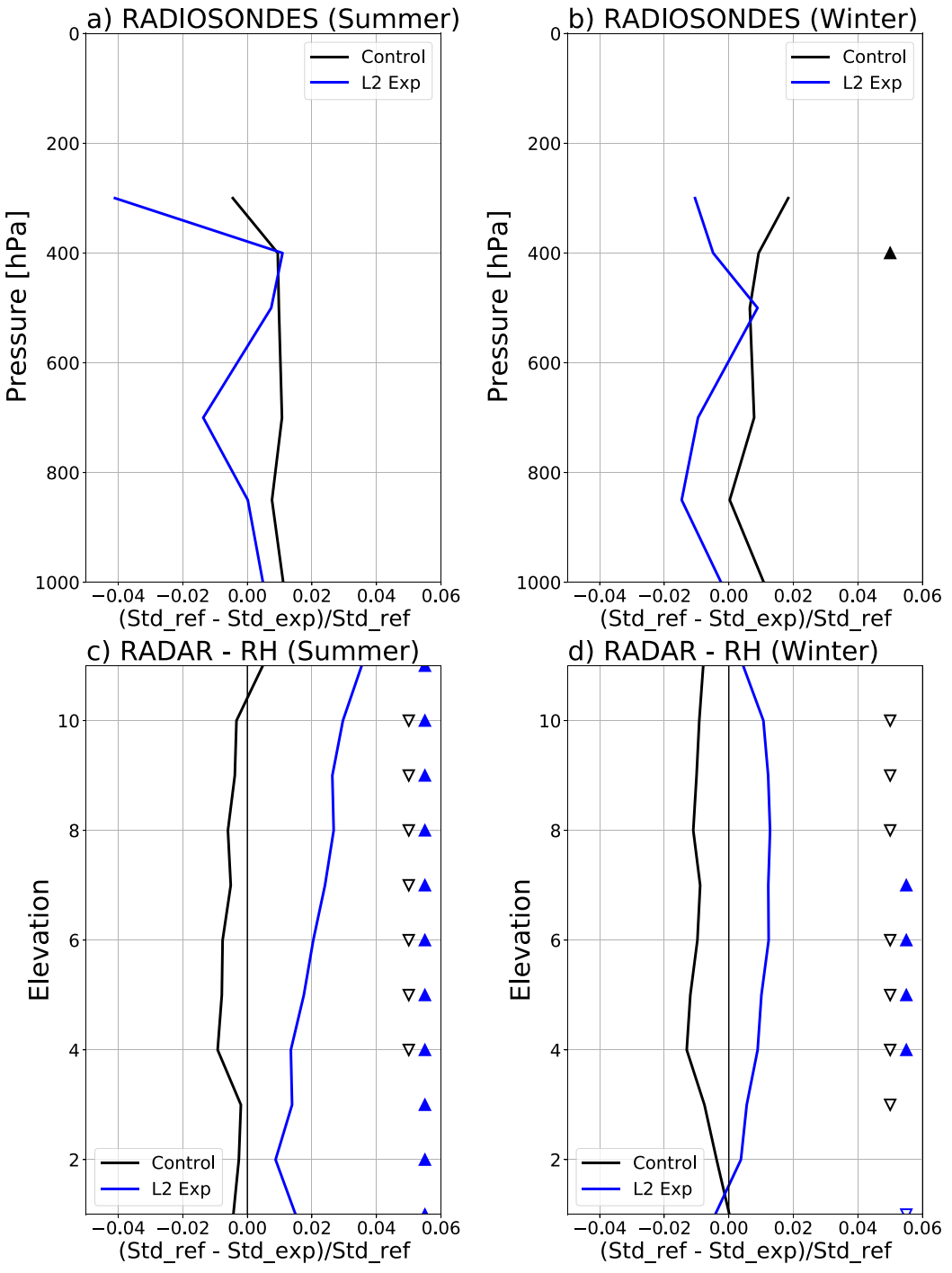

FIG. 11. As in Fig. 9, but for (a),(b) specific humidity from radiosondes and (c),(d) relative humidity (RH) retrieved from radar data.

control experiment presents a statistically significant positive impact in the relative humidity forecast above 150 hPa and in all forecast ranges during the winter (Fig. 12m). In the summertime the impact is neutral in the relative humidity forecast (Fig. 12f). The impact in the forecasts started at 1200 UTC has a similar behavior when compared to the forecast at 0000 UTC. In the 1200 UTC analyses, the L2 profiles and radiances from MetOp are assimilated; on the other hand no observation from these two datasets is assimilated at 0000 UTC. This difference can explain a higher impact at 1200 UTC.

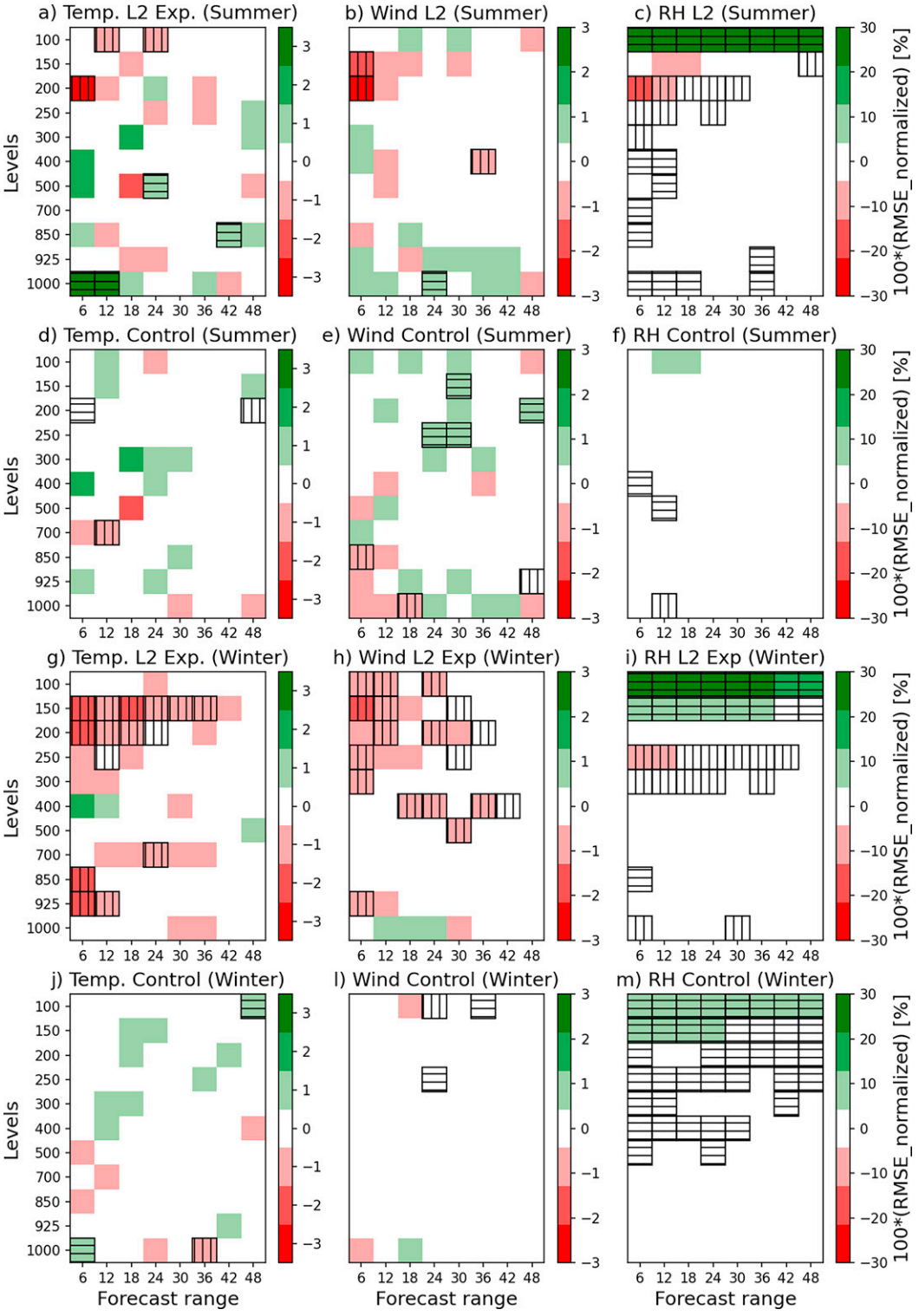

FIG. 12. Relative difference of root-mean-square error (RMSE; in %) of (a),(d),(g), (j) temperature; (b),(e),(h),(l) wind intensity; and (c),(f),(i),(m) relative humidity; 100 × [(RMSEBaseline − RMSEExp)/RMSEBaseline]. The RMSEs of baseline and experiments were calculated against ECMWF analyses. The vertical lines indicate that the baseline is better than the experiments with 95% confidence (according a Student's *t* test). Horizontal lines indicate that the experiment is better than baseline experiment with 95% confidence. Red and green colors denote a degradation and an improvement, respectively, in the experiment when compared with the baseline.

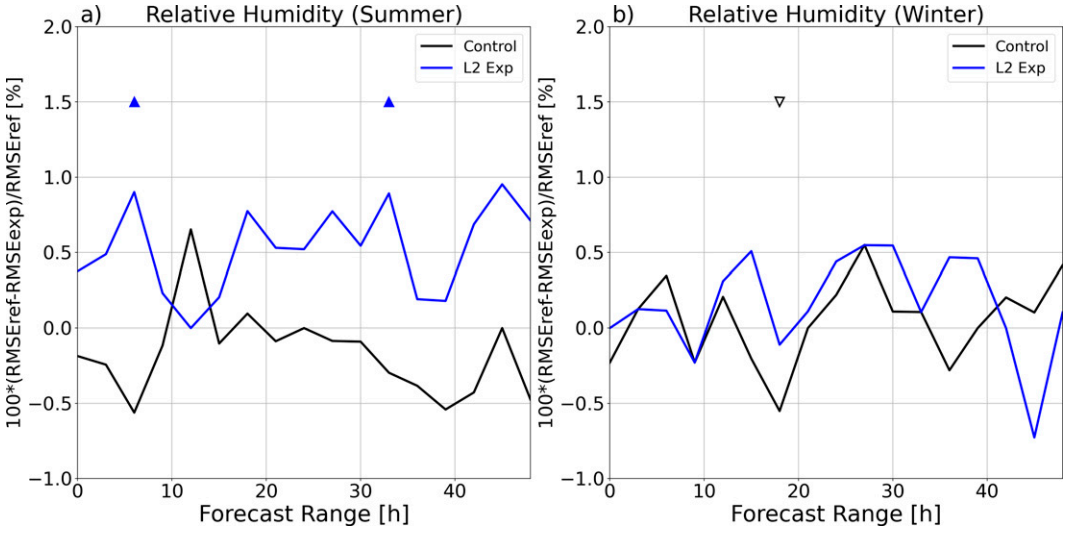

FIG. 13. As in Fig. 12, but for screen-level relative humidity evaluated against observations from surface stations (SOLFRA). (a) The summer period and (b) the winter period. The lines and triangles have the same meaning as in Fig. 11. The vertical axis is relative difference of root-mean-square error (RMSE; in %) and the horizontal axis is the forecast range.

Even though other variables were also evaluated, the ones presented above are those for which the impact in the control and L2 experiments is the greatest. The same evaluation was also made using radiosondes. The results were not shown because the L2 and control experiments presented differences that were not statistically significant, contrary to the ones shown for the evaluation using ECMWF's analyses.

2) SURFACE VERIFICATION

An evaluation was also performed against over 600 ground stations mainly located over metropolitan France. The evaluated variables were the surface pressure, 2 m temperature, 2 m relative humidity, 10 m wind, rain, and cloud cover. Only the results for the relative humidity are presented as forecast scores for the other variables are not statistically speaking significantly modified by both experiments (control and L2). This verification was performed for the forecasts started at 0000 UTC. As in the evaluation against the ECMWF's analyses, the relative differences of root-mean-square error (RMSE) in % were evaluated. The relative difference for the L2 experiment varies between 0% and 1% during summer (blue line in Fig. 13a), corresponding to a positive impact over all forecast ranges. During the winter (blue line in Fig. 13b), the relative difference varies between −0.7% and 0.5%, it is positive between 12 and 42 h of forecast. On the other hand, the control experiment presents predominantly negative impact during the summertime, except at 12 h of forecast (black line in Fig. 13b). During the winter for the control experiment, the impact mixed over the forecast ranges. In this way, a positive impact in the L2 experiment is more evident. It must be stressed that the differences between all experiments remain small.

The surface verification was also made against the synoptic observations available over the whole AROME-France domain but results are not shown because of very little difference between experiments.

3) PRECIPITATION VERIFICATION

The precipitation verification was performed for both periods against rain gauge observations. The results did not show big differences between the experiments and the baseline. Whenever differences were noticeable, they did not occur at all evaluated forecast ranges and all precipitation thresholds. For this reason, only one precipitation skill score from the summer period is chosen to represent the evaluation. This skill score effects the quality of the precipitation forecast. The chosen score (Fig. 14) is the Brier skill score with a neighborhood observation (BSS_NO; Amodei and Stein 2009). This score alleviates the problem of the double penalty and caters for mislocations of the precipitation in the forecast but with a correct intensity. The closer to 1 the score, the better the forecast. The precipitation thresholds chosen were 0.5 and 2 mm in 6 h with a neighborhood of 52.8 km. Figure 14 presents the BSS_NO of L2 (blue lines), control (black lines), and baseline (red lines). As can be seen in Figs. 14a and 14b, respectively, a degradation is found in the L2 forecast score compared to the baseline experiment at the 24 h forecast range and in the 0.5 and 2 mm precipitation thresholds. The BSS_NO of the control experiment presents an improvement in the 24 h forecast (black lines) and the differences between the baseline and the control experiments are statistically significant for the 0.5 mm threshold at 18 h (black line in Fig. 14a).

4) SUMMARY

The upper-air forecast skill score evaluation showed that temperature and wind intensity forecast scores are almost neutral in the control experiment for both periods and that

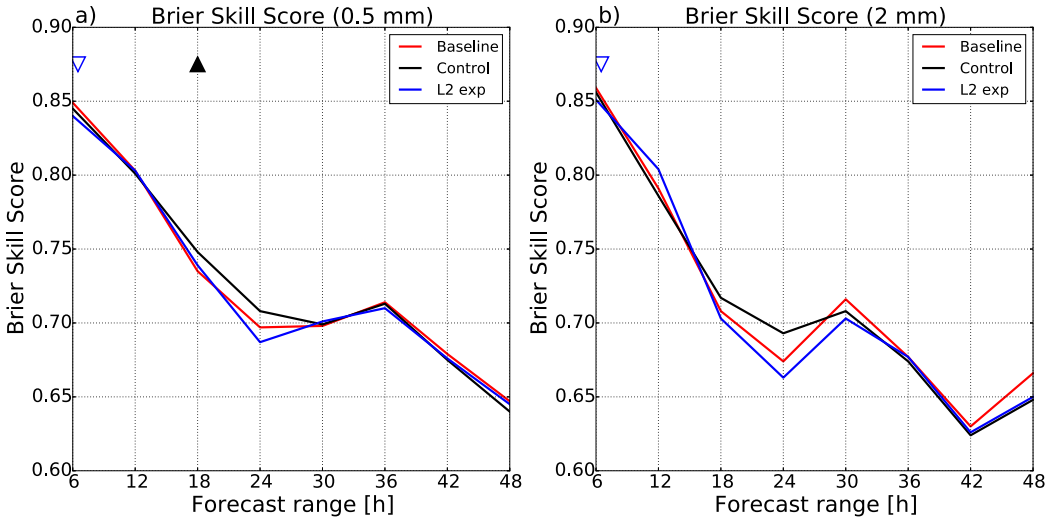

FIG. 14. Brier skill score (BSS NO) for (a) 0.5 and (b) 2 mm of precipitation accumulated in 6 h, with a neighborhood of 52.8.km. The red line is the baseline, the black line the control experiment, and the blue the L2 experiment. The plain black upward-facing triangle means that the control experiment is better than the baseline with 95% confidence (Student's *t* test). The empty black downward-facing triangle means that the baseline is better than the control experiment with 95% confidence (Student's t test). The blue triangles are similar to the black ones but are for the L2 experiment. These panels represent the summer period.

there is a slight negative value of the relative difference of RMSE at high levels in the L2 experiment. The relative humidity forecast shows a positive relative difference of RMSE in the control experiment (compared against baseline) and both positive and negative values of relative differences of RMSE in the L2 experiment when compared against the baseline. Overall, the surface skill score verification did not show any large differences but when they were any, there were positive in terms of relative humidity at 2 m for the L2 experiment. The precipitation verification scores vary during the periods analyzed. The control and L2 experiments showed an almost neutral behavior with respect to the baseline experiment.

5. Conclusions and perspectives

This work aimed to assess whether the assimilation of retrieved temperature and humidity L2 profiles, operationally produced and distributed by EUMETSAT, in the AROME-France mesoscale model instead of IASI, AMSU-A and MHS radiances was feasible and whether there would be any benefits from it. The L2 profile latency is less than 30 min from sensing, which tends to indicate that L2 profiles can be assimilated into regional models such as the AROME-France model. L2 profiles can represent a huge number of profiles to be assimilated in limited-area models. To achieve the goal of this study, first of all, the AROME-France 1-h forecasts were assessed against the L2 temperature and humidity profiles. This assessment enables the design of the configuration of the assimilation experiments using the L2 profiles instead of radiances. Last, three assimilation experiments were performed in two different periods, winter and summer, as follows: the baseline experiment, which assimilates the current set of observations in the AROME-France operational version in 2017 and 2018 except radiances from IASI, AMSU-A, and MHS sounders, the control experiment (similar to the baseline but also assimilating radiances from IASI, AMSU-A, and MHS sounders), and the L2 experiment (similar to the baseline, but also assimilating L2 profiles).

The statistics of differences between L2 and AROME-France model variables showed a good quality of L2 profiles when compared with other observation first-guess departures (radiosondes and aircraft). It is difficult to find settings to make a fair comparison between L1 and L2 assimilation and this study is a first trial that needs to be investigated further. The horizontal thinning applied to the L2 data (one profile in each 160 km box) retained the quality of the L2 profiles, which shows that the profile with the best quality indicator was rightly selected for the assimilation.

An empirical inflation factor was added to the radiosonde observation errors for temperature and humidity to estimate the L2 profiles observation error σ_{oL2} for their assimilation (i.e., 15% of relative humidity for the specific humidity profile and 1.2 times the radiosondes error profile for temperature). The observation errors for the L2 profiles (σ_{oL2}) diagnosed using the Desroziers et al. (2005) method show that the L2 profile specified observation errors (σ_{oL2}) seem to be an acceptable choice for this first study, but could be refined for a future study.

After the first statistical assessment, three assimilation experiments were defined: the baseline, the control (with L1 radiances from IASI, AMSU-A and MHS) and L2 experiments to assess the impact of L2 profiles assimilation in the AROME-France model. The configuration of the control

experiment was set up on the L2 profiles availability to provide a fair evaluation between both experiments. They ran during two periods of 2 months representing winter and summer conditions.

Results show that the L2 and control experiments helped to decrease in magnitude the mean first-guess departure (OMF) of temperature observations in the midatmosphere compared to the baseline. These findings are not observed for the standard deviation profiles, which show a degradation (at the top of the profiles). The control experiment also decreased in magnitude the mean OMF of radiosondes and aircraft temperature observations up to 300 hPa. In terms of humidity observations, the L2 experiment decreased in magnitude the radiosonde mean OMF between 800 and 700 hPa; however, at 300 hPa a degradation is observed. For relative humidity retrieved from radar observations, the mean OMF presents an improvement up to elevation 6 with a L2 profile. The standard deviation of OMF of radar observations is improved at almost all elevations. The control experiment helped to decrease the mean OMF of radiosondes between 800 and 600 hPa and up to 300 hPa. This experiment contributed, however, to the degradation of the mean OMF and standard deviation of OMF of radar. This degradation is more obvious during the winter period.

The impact of both control and L2 experiments against the baseline has also been evaluated in terms of forecast skill scores. The control experiment reduces the humidity forecast RMSE in the upper troposphere, which corresponds to the sensitivity of the eight water vapor (sensitive) assimilated IASI channels and to the four used MHS channels over land. The L2 experiment helped to reduce the humidity forecast RMSE also in the middle-lower troposphere where no L1 radiances are assimilated (assimilated water vapor channels are mainly sensitive in the high troposphere). The L2 experiment shows an increase in the humidity forecast RMSE above 300 hPa. The temperature and wind intensity of forecast RMSE of the L2 experiment were also increased at 200 hPa. The surface verification showed some differences between experiments, which are more emphasized on the bias. No clear signal can be extracted from the precipitation forecast skill during both 2-month periods. It is worth noting that IASI data represent only a small fraction of all assimilated observations (around 3% and 4.7%) in the AROME-France model and are available only 6 times out of 24 analyses made during a day.

This study shows that the L2 temperature and humidity profiles provided by EUMETSAT in near real time can be assimilated in a regional NWP model. Nevertheless, several features have to be taken into account and adjusted for each NWP system, so as to optimize the exploitation of the information of the L2 products. In particular, both horizontal and vertical data selections have to be carefully studied.

The present assimilation study is based on the operationally available EARS-IASI-L2 profiles, which are currently disseminated without their associated averaging kernels, which have therefore not been taken into account here. The fact that the impact of L2 assimilation is mixed (it depends on the atmosphere level and also on the forecast range) shows that more attention will need to be given to key aspects affecting the use of L2 products for assimilation in future L2 profile assimilation studies. Indeed, averaging kernels should make the difference in properly using the satellite constrained information. The inclusion of averaging kernels in the disseminated L2 profiles would imply a major increase of the volume of data and EUMETSAT is currently looking for ways to mitigate this problem. These efforts are based both on a lower dimensional representation of the profiles, as PC scores, and the possibility of constructing the averaging kernels at the user side from a few disseminated indices and a static lookup table disseminated only once.

In this study, the assimilation experiments were carried out in a simplified manner, taking advantage of the existing assimilation framework for radiosondes in AROME-France. The L2 profiles were therefore assumed as pseudoradiosondes, i.e., without considering vertical error correlation and the vertical sensitivity. On the latter point, as the quality within a profile varies with the occurrence of clouds and its top pressure within a pixel, additional information, e.g., such as cloud top or uncertainty along the profile, could be useful for determining which parts of the profiles would be suitable for assimilation. An evaluation shows for one analysis that the L2 profiles are assimilated preferentially in clear sky and low cloud areas. In the current setting, the quality indicator is pixel-based and a whole profile can be rejected by the assimilation system because the quality indicator is a single uncertainty estimate relating to the profile quality in the lower troposphere. In preparation for the IRS launch, a specific assimilation study shall also be performed using Level 2 profiles with infrared observations only, but not merged with microwave data. In this configuration, a detailed cloud characterization or uncertainty estimates on the vertical would be of the utmost importance.

Acknowledgments. The authors have no conflict of interest to disclose. This research has received funding from the EUMETSAT study (EUM/CO/17/4600001975/TA).

Data availability statement. The source code of AROME-France used in this paper is not freely available, but the simulation data that support the findings of this study are available from the corresponding author on reasonable request.

REFERENCES

Amodei, M., and J. Stein, 2009: Deterministic and fuzzy verification methods for a hierarchy of numerical models. *Meteor. Appl.*, **16**, 191–203, https://doi.org/10.1002/met.101.

Auligné, T., A. P. McNally, and D. P. Dee, 2007: Adaptive bias correction for satellite data in a numerical weather prediction system. *Quart. J. Roy. Meteor. Soc.*, **133**, 631–642, https://doi.org/10.1002/qj.56.

Brousseau, P., Y. Seity, D. Ricard, and J. Léger, 2016: Improvement of the forecast of convective activity from the AROME-France system. *Quart. J. Roy. Meteor. Soc.*, **142**, 2231–2243, https://doi.org/10.1002/qj.2822.

Coniglio, M. C., G. S. Romine, D. D. Turner, and R. D. Torn, 2019: Impacts of targeted AERI and Doppler lidar wind retrievals on short-term forecasts of the initiation and early evolution of thunderstorms. *Mon. Wea. Rev.*, **147**, 1149–1170, https://doi.org/10.1175/MWR-D-18-0351.1.

Courtier, P., C. Freydier, J.-F. Geleyn, F. Rabier, and M. Rochas, 1991: The Arpege project at Meteo France. *Seminar on Numerical Methods in Atmospheric Models,* Vol. II, Shinfield Park, Reading, ECMWF, 193–232, https://www.ecmwf.int/node/8798.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396, https://doi.org/10.1256/qj.05.108.

EUMETSAT, 2017a: IASI L2 PPF v6.3 validation report. EUMETSAT, 45 pp., https://www.eumetsat.int/media/45996.

——, 2017b: IASI level 2: Product guide. EUMETSAT, 80 pp., https://www.eumetsat.int/media/45982.

——, 2018: IASI L2 PPF v6.4 validation report. EUMETSAT, 59 pp., https://www.eumetsat.int/media/45739.

Eyre, J. R., S. J. English, and M. Forsythe, 2019: Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quart. J. Roy. Meteor. Soc.*, **146**, 49–68, https://doi.org/10.1002/qj.3654.

Feltz, M., R. Knuteson, S. Ackerman, and H. Revercomb, 2014: Application of GPS radio occultation to the assessment of temperature profile retrievals from microwave and infrared sounders. *Atmos. Meas. Tech.*, **7**, 3751–3762, https://doi.org/10.5194/amt-7-3751-2014.

Guidard, V., N. Fourrié, P. Brousseau, and F. Rabier, 2011: Impact of IASI assimilation at global and convective scales and challenges for the assimilation of cloudy scenes. *Quart. J. Roy. Meteor. Soc.*, **137**, 1975–1987, https://doi.org/10.1002/qj.928.

Hartung, D. C., J. A. Otkin, R. A. Petersen, D. D. Turner, and W. F. Feltz, 2011: Assimilation of surface-based boundary layer profiler observations during a cool-season weather event using an observing system simulation experiment. Part II: Forecast assessment. *Mon. Wea. Rev.*, **139**, 2327–2346, https://doi.org/10.1175/2011MWR3623.1.

Hilton, F., and Coauthors, 2012: Hyperspectral Earth Observation from IASI: Four years of accomplishments. *Bull. Amer. Meteor. Soc.*, **93**, 347–370, https://doi.org/10.1175/BAMS-D-11-00027.1.

Hu, J., N. Yussouf, D. D. Turner, T. A. Jones, and X. Wang, 2019: Impact of ground-based remote sensing boundary layer observations on short-term probabilistic forecasts of a tornadic supercell event. *Wea. Forecasting*, **34**, 1453–1476, https://doi.org/10.1175/WAF-D-18-0200.1.

Migliorini, S., 2012: On the equivalence between radiance and retrieval assimilation. *Mon. Wea. Rev.*, **140**, 258–265, https://doi.org/10.1175/MWR-D-10-05047.1.

Otkin, J. A., D. C. Hartung, D. D. Turner, R. A. Petersen, W. F. Feltz, and E. Janzon, 2011: Assimilation of surface-based boundary layer profiler observations during a cool-season weather event using an observing system simulation experiment. Part I: Analysis impact. *Mon. Wea. Rev.*, **139**, 2309–2326, https://doi.org/10.1175/2011MWR3622.1.

Rodgers, C., 2000: *Inverse Methods for Atmospheric Sounding: Theory and Practice.* Series on Atmospheric, Oceanic and Planetary Physics, Vol. 2, World Scientific, 256 pp.

Roman, J., R. Knuteson, T. August, T. Hultberg, S. Ackerman, and H. Revercomb, 2016: A global assessment of NASA AIRS v6 and EUMETSAT IASI v6 precipitable water vapor using ground-based GPS SuomiNet stations. *J. Geophys. Res. Atmos.*, **121**, 8925–8948, https://doi.org/10.1002/2016JD024806.

Salonen, K., T. August, T. Hulberg, and A. McNallya, 2019: Impact assessment of IASI temperature and humidity retrievals in the ECMWF system. *Int. TOVS Study Conf.*, Saint-Sauver, Québec, Canada, ECMWF/EUMETSAT, https://cimss.ssec.wisc.edu/itwg/itsc/itsc22/posters/7p.01.salonen.pdf.

Saunders, R., and Coauthors, 2018: An update on the RTTOV fast radiative transfer model (currently at version 12). *Geosci. Model Dev.*, **11**, 2717–2737, https://doi.org/10.5194/gmd-11-2717-2018.

Seity, Y., P. Brousseau, S. Malardel, G. Hello, P. Bénard, F. Bouttier, C. Lac, and V. Masson, 2011: The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991, https://doi.org/10.1175/2010MWR3425.1.

Wattrelot, E., O. Caumont, and J.-F. Mahfouf, 2014: Operational implementation of the 1D+3D-Var assimilation method of radar reflectivity data in the AROME model. *Mon. Wea. Rev.*, **142**, 1852–1873, https://doi.org/10.1175/MWR-D-13-00230.1.