Global Observing System Experiments within the Météo-France 4D-Var Data Assimilation System

P. Chambon, J.-F. Mahfouf, O. Audouin, C. Birman, N. Fourrié, C. Loo, M. Martet, P. Moll, C. Payan, V. Pourret, et al.

Global observing system experiments within the Météo-France 4D-Var data assimilation system

P. Chambon,[a] J.-F. Mahfouf,[a] O. Audouin,[a] C. Birman,[a] N. Fourrié, [a] C. Loo,[a] M. Martet,[a] P. Moll, [a] C. Payan, [a] V. Pourret, [a] and D. Raspaud [a]

[a] *CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France*

Corresponding author: J.-F. Mahfouf, jean-francois.mahfouf@meteo.fr

ABSTRACT: Observing System Experiments were undertaken within the 4D-Var data assimilation of the Météo-France global Numerical Weather Prediction (NWP) model. A six-month period was chosen (October 2019 - March 2020) where 40 millions of observations per day were assimilated. The importance of in-situ observations provided by aircraft, radiosondes and surface weather stations, despite their small fractional amount (7 %), has been confirmed particularly in the Northern Hemisphere. Moreover, the largest impact over Europe in terms of Root Mean Square Error (RMSE) scores comes from surface observations. Satellite data play a dominant role over tropical regions and the Southern Hemisphere. Microwave radiances have a more pronounced impact on the long range and on the humidity field than infrared radiances, despite being less numerous (10 % versus 80 %). Bending angles impact significantly the quality of the upper troposphere / lower stratosphere temperature of the tropics and Southern Hemisphere. Atmospheric Motion Vectors (AMVs) are beneficial in wind forecasts at low and high levels in the tropics and the Southern Hemisphere, but also in the humidity field. Such impacts are only significant during the first 48 hours of the forecasts. Scatterometer winds have an impact restricted to low levels which is kept at longer ranges. A comparison with Forecast Sensitivity - Observation Impact studies over a 3 month period using the same measure of short-range (24 h) forecast errors reveals that the ranking between the major observing systems is kept between these two ways of measuring observation impact in NWP. From our conclusions, recommendations are provided on possible evolutions of the global observing system for NWP.

1. Introduction

The forecast skill of Numerical Weather Prediction (NWP) models has steadily improved during past decades due to a more efficient usage of satellite observations within advanced data assimilation systems, such as four-dimensional variational (4D-Var) schemes (Simmons and Hollingworth 2001). These improvements are also the result of rapid technological developments in the field of High Performance Computers (HPCs). Indeed, with more powerful HPCs it has been possible to use NWP models at higher spatial resolutions with more accurate numerical and physical process representations.

Within national weather services, operational NWP upgrades are often the result of many contributions: changes to the observing systems, increases in horizontal and/or vertical resolutions, revisions to the numerical and physical processes, and more members in ensemble systems (for prediction and assimilation). In order therefore to isolate the contribution of changes in terms of observation usage, it is necessary to perform dedicated sensitivity experiments, which are known as *Observing System Experiments (OSEs)* within which a specific observing system is withdrawn from a baseline comprehensive system (e.g. Radnóti et al. (2012)). It is important to regularly assess the value of observations in a NWP data assimilation context, for data producers (to justify the maintenance and the evolution of observing networks and satellite programs given the associated costs), for an improved usage (when the withdrawal of observations leads to unexpected improved scores) and to evaluate the robustness of the data assimilation system (to identify the most sensitive observing systems that may require consolidation).

Since the World Meteorological Organization (WMO) provides recommendations to data producers in order to maintain a comprehensive global observing network, regular workshops are organized in order to review the observation data usage in NWP models with results from OSEs. In the context of the 7th WMO impact workshop [1](30 November - 3 December 2020), Météo-France has performed a number of OSEs with a recent version of their global NWP model to be described in this paper. In Section 2, a reference system is described (the main features of the global NWP model and its data assimilation system) with a baseline observing system corresponding to the one used operationally during the first half of the year 2020. The experimental design is explained in Section 3 (period of interest and the set of observation denial experiments). The main results are

[1]https://community.wmo.int/meetings/NWP-7

presented in Section 4 in terms of short- and medium-range forecast skill scores. For an improved understanding of these results, additional denial experiments have been performed, and the main outcomes are described in Section 5. In this Section a comparison of OSEs results with those obtained from the Forecast Sensitivity Observation Impacts (FSOI) adjoint method (Langland and Baker 2004) is shown. The main conclusions of the study are summarized in Section 6, including a number of recommendations on observation usage in the NWP context.

2. Description of the reference system

a. The NWP model

The global spectral NWP model ARPEGE (Action de Recherche Petite Echelle Grande Echelle), based on a numerical code jointly developed between Météo-France and the European Centre for Medium-range Weather Forecasts (ECMWF), is used in this study (Courtier et al. 1991). An original feature of this model is its tilted and stretched conformal horizontal grid (Courtier and Geleyn 1988) which allows an increased resolution over Europe (the region of main interest for numerical forecasts run up to 4 days at Météo-France). The current operational system (CY43T2 between July 2019 and June 2022) has a spectral resolution T_L1798 (triangular truncation up to wave number 1798 associated to a linear reduced Gaussian grid). The stretching factor of the transform grid ($c = 2.2$) leads to a horizontal resolution of about 5 km over Europe and 25 km at the antipodes of the numerical pole (around New Zealand). In the context of the current OSEs, we have chosen this model cycle but with a coarser horizontal resolution. This will allow experiments to be conducted over longer periods of time in order to increase the significance of the differences and also to consider a larger number of scenarios. This choice has been guided by ECMWF experience in this context (McNally 2012; Bormann et al. 2019). The selected truncation is T_L798, which was used operationally in Météo-France from 2010 to 2015, corresponds to a resolution of 11 km over Europe and 55 km at the antipodes. The vertical grid is discretized in 105 levels with a hybrid pressure terrain-following coordinate system η from 10 m above ground up to 0.01 hPa. Additional details on the ARPEGE model regarding the prognostic equations, their numerical resolution and the physical parameterization schemes can be found in Bouyssel et al. (2022).

b. The 4D-Var assimilation system

The initial conditions of the ARPEGE model are provided by a 4D-Var data assimilation system with a 6-hour assimilation window and 30-min observation time-slots. The incremental formulation proposed by Courtier et al. (1994) solves the minimization of a quadratic cost-function expressed in terms of increments at coarser resolution with trajectory updates (so-called "outer-loops"). In the operational context the first minimization is performed at truncation T_L224 (of around 100 km) whereas the second one uses a higher truncation T_L499 (of around 40 km). A set of 40 iterations is chosen for each minimization as a compromise between the computing time and the convergence of the cost-function. In order to make the 4D-Var more efficient, this operational set-up has been modified for the OSEs where the second minimisation uses the same truncation as the first one. In terms of linearized physical parameterizations, the first minimization includes only a vertical diffusion scheme (neglecting perturbations of exchange coefficients) whereas the second one accounts also for large-scale condensation and gravity wave drag schemes. A dedicated surface analysis based on optimal interpolation schemes is performed every 6 hours (central time of the 4D-Var window) over oceans (sea surface temperature) and continents (screen-level temperature and relative humidity; moisture content and temperature in the soil) using in-situ measurements from SYNOP, BUOY and SHIP reports.

An Ensemble Data Assimilation (EDA) is coupled to the 4D-Var system in order to provide flow dependent background error covariances. The ensemble is made of 50 members using a low resolution and a simplified 4D-Var configuration (one outer-loop) compared to the deterministic run. The EDA allows the estimation of background error standard deviations and correlations lengths of the variables to be initialized in a wavelet block-diagonal formulation of the correlation matrix. Additional details are provided in Bouyssel et al. (2022). The OSEs will consider the EDA background errors from the operational system even though it is known that changing the observing system modifies background errors. Indeed a reduced (resp. enhanced) observing system is expected to decrease (resp. increase) the quality of the forecast leading to larger (resp. smaller) background errors. Such property has been exploited to assess the impact of observing systems in a NWP context (Tan et al. 2007; Harnisch et al. 2013). The computational cost of rerunning the EDA would prevent us however from performing a large set of experiments. This common practice

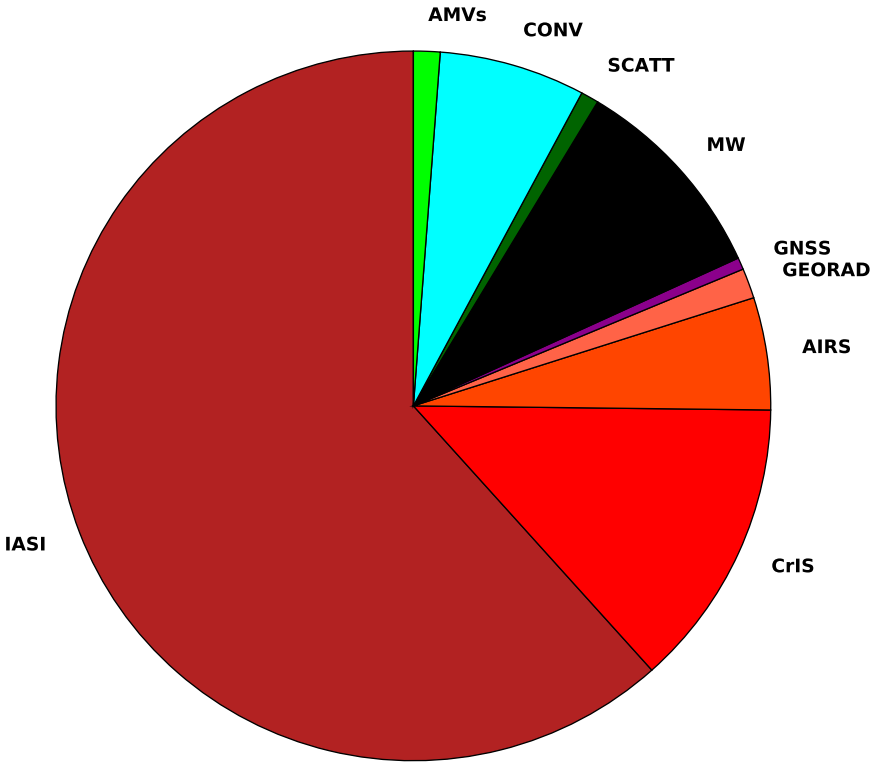


FIG. 1. Main observation types assimilated in the ARPEGE 4D-Var assimilation system over the period October 2019 to March 2020 and described more precisely in Table 1. The infrared radiances from polar orbiting satellites are shown in red (IASI, CrIS, AIRS), those from geostationary satellites (GEORAD) in orange, the AMVs in green, the in-situ observations (CONV) in cyan, the oceanic surface winds from scatterometers (SCATT) in olive, the microwave radiances (MW) in black and the GNSS-RO bending angles (GNSS) in purple.

has been recently confirmed by OSE results from Duncan et al. (2021) who showed that the effects of updating background errors is secondary to that caused by the observing-system change itself.

c. The baseline observing system

Table 1 summarizes the baseline observing system chosen in the reference 4D-Var assimilation. It corresponds to the set of observations assimilated operationally in the ARPEGE model at Météo-France from January to July 2020. Indeed in January 2020, a last instrument from *Metop-C* was introduced (ASCAT[2] on top of AMSU-A[3], MHS[4], GRAS[5] and IASI[6]) whereas in July 2020 the constellation of GNSS-RO[7] receivers was considerably enhanced (9 new instruments) and winds from Aeolus lidar added (not considered here). The availability of 3 Metop satellites and two recent NOAA platforms (S-NPP, NOAA-20) has allowed the ARPEGE model to assimilate around 40 millions of observations per day (this may be considered as a *golden age* for NWP models in terms of data availability). With six hyperspectral infrared sounders (3 IASI, 2 CrIS[8] and 1 AIRS[9]) the observing system is dominated by their radiances which represent 80 % of the total observations (Figure 1). With 18 radiometers, microwave radiances reach a fractional amount of 10 %. Other spaceborne instruments represent less than 3 % (GNSS-RO, AMVs[10], scatterometer winds), whereas the percentage of in-situ conventional data (aircraft, sondes, surface stations) is only 7 %. In order to avoid spatial observation error correlations, most satellite data are thinned at 140 km. This distance is increased to 280 km for AMVs and reduced to 100 km for IASI radiances and scatterometer winds. Interchannel correlations errors are specified for the hyperspectral infra-red sounders IASI and CrIS from a-posteriori diagnostics (Desroziers et al. 2005). These correlations are currently neglected for other satellite radiances. Satellite radiance biases are identified in the 4D-Var system using a variational bias correction technique with suitable predictors (Auligné et al. 2007). Regarding surface observations for the upper air analysis, surface pressure observations from SYNOP (over land), SHIP and BUOY reports (over oceans) are assimilated in terms of geopotential height. Oceanic surface winds from SHIP reports and relative humidity from SYNOP reports (during daytime only) are also used.

The geographical distribution of the main observing systems examined hereafter are displayed in Figure 2 for a 6-hour period corresponding to the length of the 4D-Var assimilation window.

[2]ASCAT: Advanced Scatterometer
[3]AMSU-A: Advanced Microwave Sounding Unit-A
[4]MHS: Microwave Humidity Sounder
[5]GRAS: Global Navigation Satellite System Receiver for Atmospheric Sounding
[6]IASI: Infrared Atmospheric Sounding Interferometer
[7]GNSS-RO: Global Navigation Satellite System - Radio Occultation
[8]CrIS: Cross-track Infrared Sounder
[9]AIRS: Atmospheric Infrared Sounder
[10]AMVs: Atmospheric Motion Vectors

Surface observations have the highest density over Europe. There is good coverage over Asia, the Americas and Australia. On the other hand the number of stations is very much reduced over Africa. The radiosonde network exhibits a hemispheric disparity with a good coverage over Europe, North America, Russia and China, and poor coverage over the Southern Hemisphre. There are few stations over the tropics and in the Southern Hemisphere, due to the presence of oceans and to continental data voids in South America and Africa. In terms of aircraft data, the highest density is over North America and Europe, and between the two continents. Regional commercial airlines can be seen over Europe, China and Australia. Similarly to other in-situ observations, the Southern Hemisphere lacks aircraft data. The amount of polar orbiting microwave radiometers on contrasted orbits allows a global coverage over a 6-hour period. When considering hyperspectral infrared sounders, despite representing the largest data amount, the coverage is not complete over 6 hours, because of only two complementary orbits for the polar satellites. The coverage of AMVs is important between 50°S and 50°N. Polar orbiting satellites provide additional wind information at high latitudes between 70° and 90°. The coverage provided by scatterometers for oceanic surface wind is far from optimal since the three Metop satellites are on the same orbit and there is only one additional satellite (OSCAT on Scatsat-1). This statement is also true for the GNSS-RO bending angles because only two complementary orbits are available in addition to the Metop satellites.

3. Experimental design

All experiments were run over a 6-month period from October 1, 2019 to March 31, 2020. This period has been chosen since, as explained above, it is associated with a wealth of satellite observations. The data latency windows for observation usage in 4D-Var were taken from the operational system with values ranging from 70 to 180 min depending upon analysis time. From each analysis at 00 UTC (the background field being a 6-h forecast that starts at 18 UTC the day before), a 4-day forecast model integration at resolution T_L798 was run and compared against radiosoundings and ECMWF operational analyses (assumed to be independent measures of the true state of the atmosphere).

We have considered a baseline experiment (**REF**) with the full observing system. It has been verified that the quality of the resulting analyses and forecasts is rather similar to the one from the operational system, despite slightly lower objective skill scores (in terms of RMSE values due

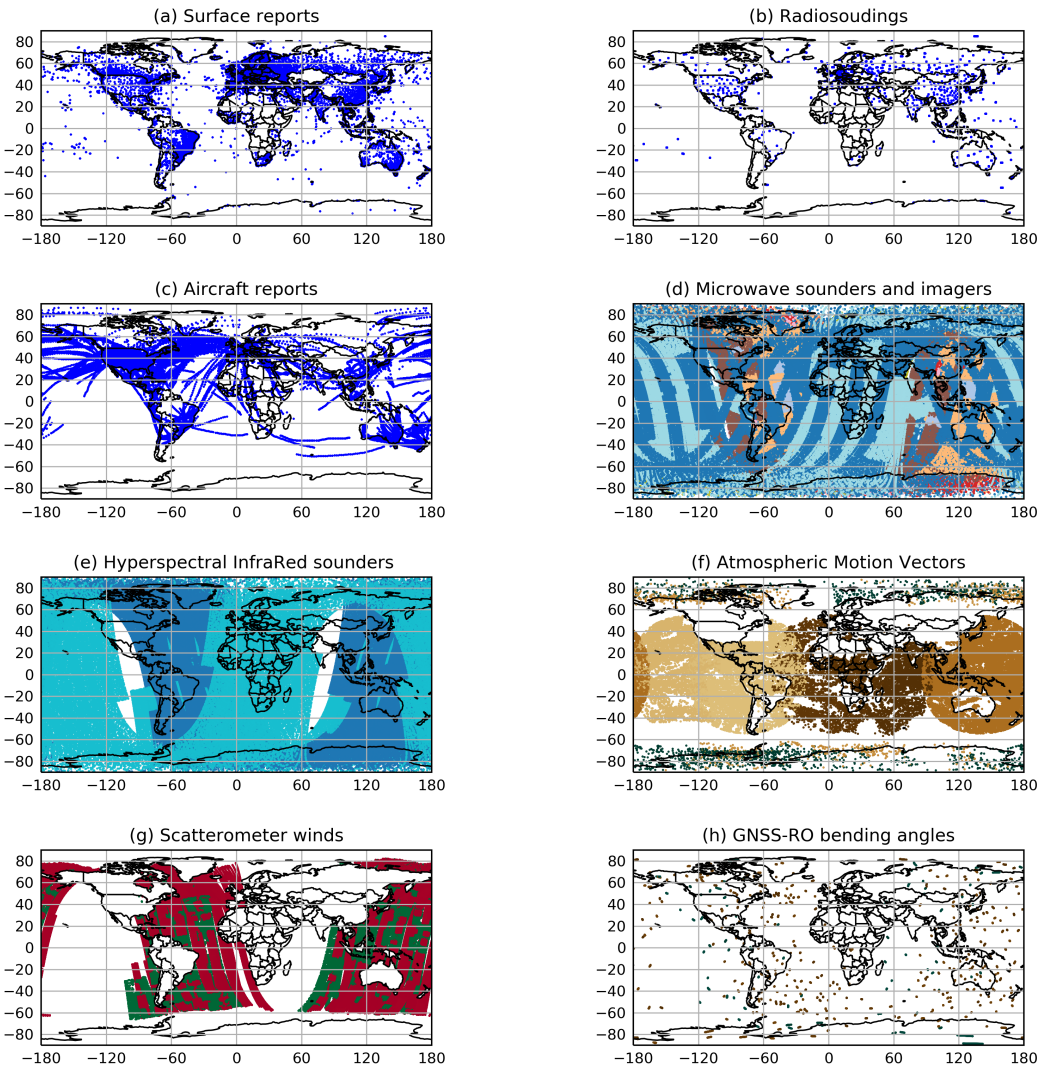


FIG. 2. Geographical coverage of the main observing systems evaluated in the OSEs. a) surface stations, b) radiosounding stations, c) aircraft data, d) microwave radiances, e) hyperspectral infrared radiances, f) atmospheric motion vectors, g) scatterometer winds, h) GNSS-RO bending angles for a 6-h period around 01/10/2019 at 00 UTC. The various colors allow to distinguish different satellite platforms for a particular observing system or observation type.

TABLE 1. Summary of the observing systems assimilated in the baseline 4D-Var system of the ARPEGE model. The maximum number of radiances per spaceborne sensor is provided in the last column with their sensitivity to temperature (T), water vapor (WV) and ozone (O_3). Similary, the spectral bands (VIS, IR, WV) used for the derivation of atmospheric wind vectors are shown. In-situ sensors measure suface pressure (P_s), temperature (T), relative humidity (RH) and winds. The ground based GNSS (GB-GNSS) receivers provide Zenith Total Delays (ZTD) measurements informative on integrated water vapor.

Observation type	Instruments / Platform	Comments
LEO IR	IASI (Metop-A/B/C)	129 channels (T, WV, O_3)
radiances	CrIS (S-NPP, NOAA-20)	68 channels (T, WV)
	AIRS (Aqua)	75 channels (T)
GEO IR	SEVIRI (Meteosat-8/11)	6 channels (T, WV)
radiances	AHI (Himawari-8)	5 channels (T, WV)
LEO MW	AMSU-A (NOAA-15/18/19, Aqua, Metop-A/B/C)	9 channels (T)
radiances	ATMS (S-NPP, NOAA-20)	14 channels (T, WV)
	MHS (NOAA-19, Metop-A/B/C)	3 channels (WV)
	MWHS-2 (FY-3C)	3 channels (WV)
	SAPHIR (Megha-Tropiques)	6 channels (WV)
	SSMI/S (DMSP F-17/18)	14 channels (T, WV)
	GMI (GPM-Core)	2 channels (WV)
GNSS-RO	GRAS (Metop-A/B/C)	above 8 km
bending angles	IGOR (COSMIC-1)	-
	IGOR (TerraSAR-X)	-
	IGOR (TanDEM-X)	-
Scatterometer	C-band ASCAT (Metop-A/B/C)	neutral 10-m winds
surface winds	Ku-band OSCAT (ScatSat-1)	neutral 10-m winds
AMVs	SEVIRI (Meteosat-8/11)	(WV, IR, VIS)
	ABI (GOES-15, 16)	(WV, IR, VIS)
	AHI (Himawari-8)	(WV, IR, VIS)
	MODIS (Terra, Aqua)	(WV, IR)
	AVHRR (NOAA-15, 18, 19)	(IR)
Aircrafts	AIREP, AMDAR	(T, winds)
Sondes	PILOT, TEMP, Profilers	(T, RH, winds)
Surface	BUOY, SHIP, SYNOP, GB-GNSS	(P_s, T, RH, winds, ZTD)

to the coarser horizontal resolution, both in the data assimilation system and in the model (not shown).

A set of 6 denial experiments excluding the following observing systems was then undertaken:

- **NO CONV**: no in-situ conventional observations (radiosoundings, aircraft reports, wind profilers, SYNOP stations, SHIP and BUOY reports)
- **NO MW**: no microwave radiances from imaging and sounding radiometers (18 instruments)
- **NO IR**: no infra-red radiometers from polar orbiting satellites (6 hyperspectral sounders) and geostationary satellites (3 imagers)
- **NO AMVs**: no Atmospheric Motion Vectors from polar orbiting (5 platforms) and geostationary satellites (4 platforms)
- **NO GNSS**: no bending angles from low-orbiting GNSS-RO receivers (6 instruments)
- **NO SCATT**: no ocean surface winds from scatterometers (4 instruments)

Note that in all the above experiments, observations used to produce the surface analyses have not been modified. It is also worth mentioning that each experiment has its own variational bias correction scheme within the 4D-Var system allowing possible changes induced by the reduced obervational datasets.

4. Main results

a. Short-range impacts

It has been observed that for all denial experiments there is a better fit of the analysis state to the remaining observing systems, but that, on the other hand, the fit of the background state (6-h forecast) to the remaining observations is generally degraded. Satellite radiance biases do not appear to be particularly increased in both the experiments where the so-called "anchoring data" (Eyre 2016) are removed: **NO CONV** and **NO GNSS**. It is likely that the role taken by one of these is enhanced in the experiment where the other one is removed. Such behavior can reassure by showing the robustness of the current observing system, thanks to some redundancy. The standard deviation of background departures normalized by **REF** are displayed in Figure 3 against radiosoundings.

Regarding temperature, the largest degradation reaching 10 % takes place in the Southern Hemisphere near 200 hPa from **NO GNSS**. This experiment leads to similar degradations in the Northern Hemisphere near 100 hPa but with smaller values (between 1.5 and 2 %). This can

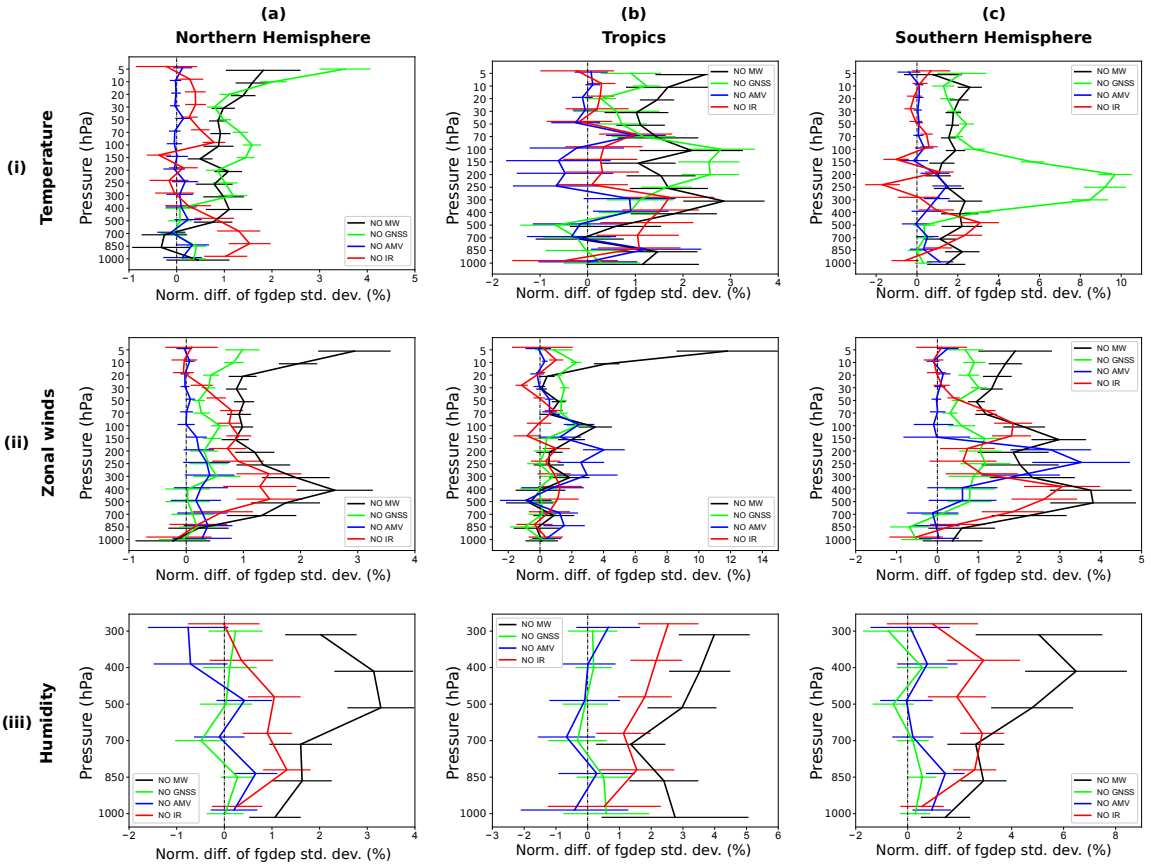


FIG. 3. Standard deviation of background departures, normalized by the reference **REF**, for the Northern Hemisphere extra-tropics above latitude 20°N (left), the tropics between latitudes 20°S and 20°N (middle), and the Southern Hemisphere extra-tropics below latitude 20°S (right). The observations are temperature from radiosondes (top), vector wind from radiosondes (middle) and specific humidity from radiosondes (bottom). Statistics cover the period October 2019 to March 2020 (6 months). Positive values indicate an increase in the background error due to the denial of the respective observing system (**NO MW**, **NO GNSS**, **NO AMVs**, **NO IR**). Horizontal lines indicate the 99 % level of statistical significance.

be explained by the fact that radiosoundings (providing vertical profiles of temperature up to 30 km) and aircraft data (providing temperature information at cruise level near 10 km) are very few in the Southern Hemisphere with respect to the Northern Hemisphere (as clearly displayed in Figure 2). Microwave instruments have a short-range impact of around 1.5 % over the whole troposphere and in the stratosphere. This impact reaches the surface in the Southern Hemisphere, but it is less significant at low levels in the tropics, and appears to be slighly negative over the Northern Hemisphere (which could be the signature of a non-optimal usage of surface sensitive

channels over continents and/or sea-ice). These negative impacts are likely more pronounced over the Northern Hemisphere (corresponding to the winter period) due to larger sea-ice extents and to the presence of clouds which could affect the surface emissivity retrieval based on the method of Karbou et al. (2014). Concerning infra-red radiances, their important contribution is in the low and mid-troposphere (up to 1.5 % in the Northern Hemisphere). A small significant impact (0.5 %) is noticed in the stratosphere (above 70 hPa) of the Northern Hemisphere. As expected, the impact of **NO AMVs** is rather weak on the temperature field but there is a small significant effect over the tropics of around 300 hPa and in the Southern Hemisphere of around 200 hPa.

On the zonal wind, the experiment **NO GNSS** is the one which has the lowest impact but with small detrimental effects (i.e. positive values) above 300 hPa in the extra-tropics and above 100 hPa in the tropics (between 0.5 and 1 %). Microwave and infra-red radiances represent the major extra-tropical contribution in the mid-troposphere with a dominant effect of **NO MW** in the stratosphere above 50 hPa. Such indirect impact is a consequence of both the multi-variate background error covariance matrix and the explicit model dynamics used to fit observations at the appropriate time in a 4D-Var system. In the tropics and in the Southern Hemisphere, the experiment **NO AMVs** degrades the 6-h forecast by up to 3 % around 200 hPa (demonstrating the importance of winds deduced from high level cloud motions).

Specific humidity reveals that the most important contribution is provided by the microwave instruments leading up to 6 % degradation in the upper troposphere of the Southern Hemisphere. The impact of infra-red sounders is smaller by a factor of 3 in the extra-tropics and by a factor of 2 in the tropics. The **NO AMVs** and **NO GNSS** experiments do not significantly impact atmospheric humidity as these two observing systems are not directly sensitive to this quantity . A small impact is noticed below 850 hPa for the **NO AMVs** experiment over extra tropical regions which could be explained by advection processes.

These results appear to be consistent with those presented by Bormann et al. (2019) with the ECMWF 4D-Var system. A larger impact of infra-red sounders observed in our experiments is likely due to the fact that our baseline system includes 6 hyperspectral instruments compared to only four at ECMWF.

b. Medium-range impacts

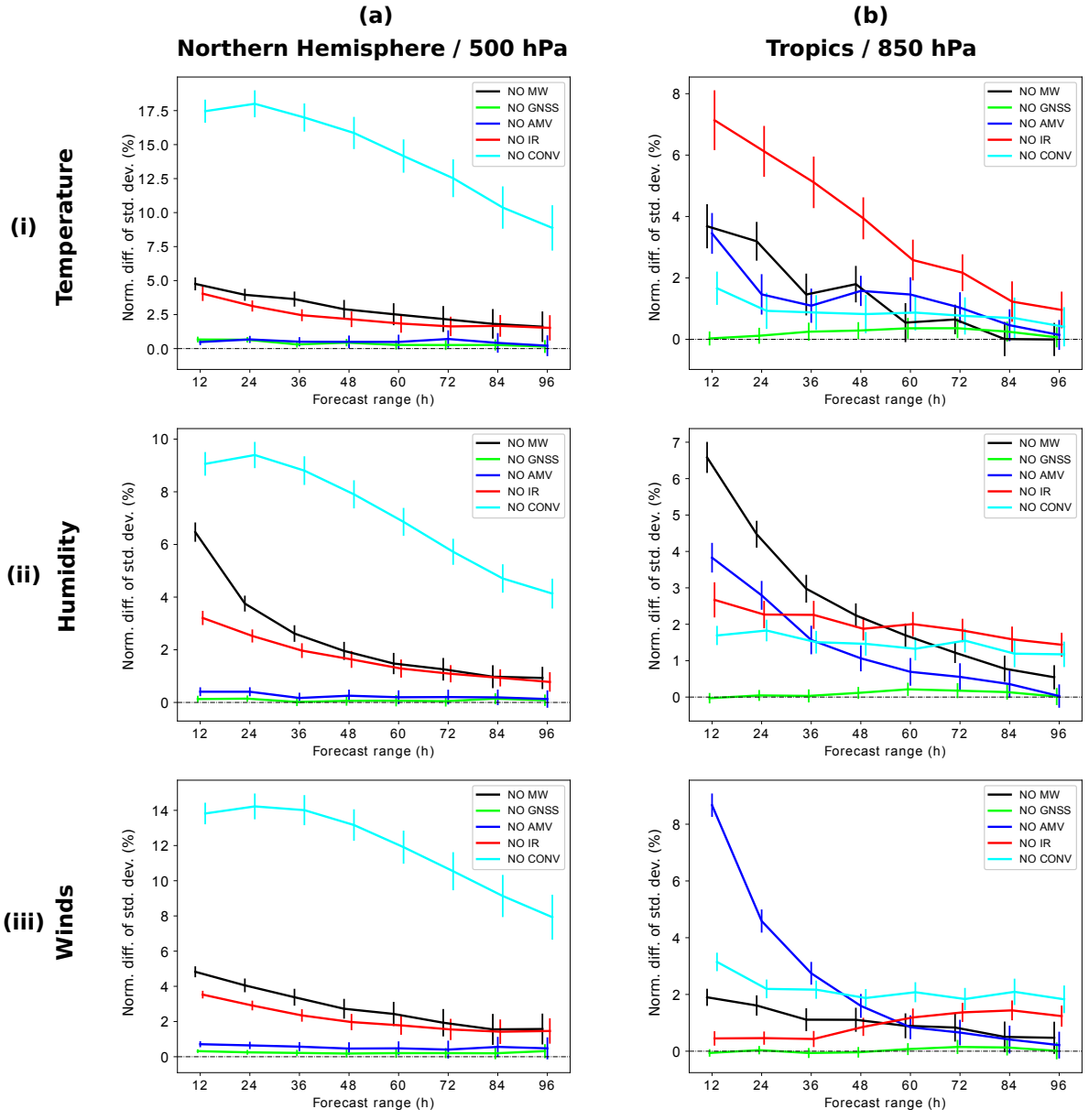
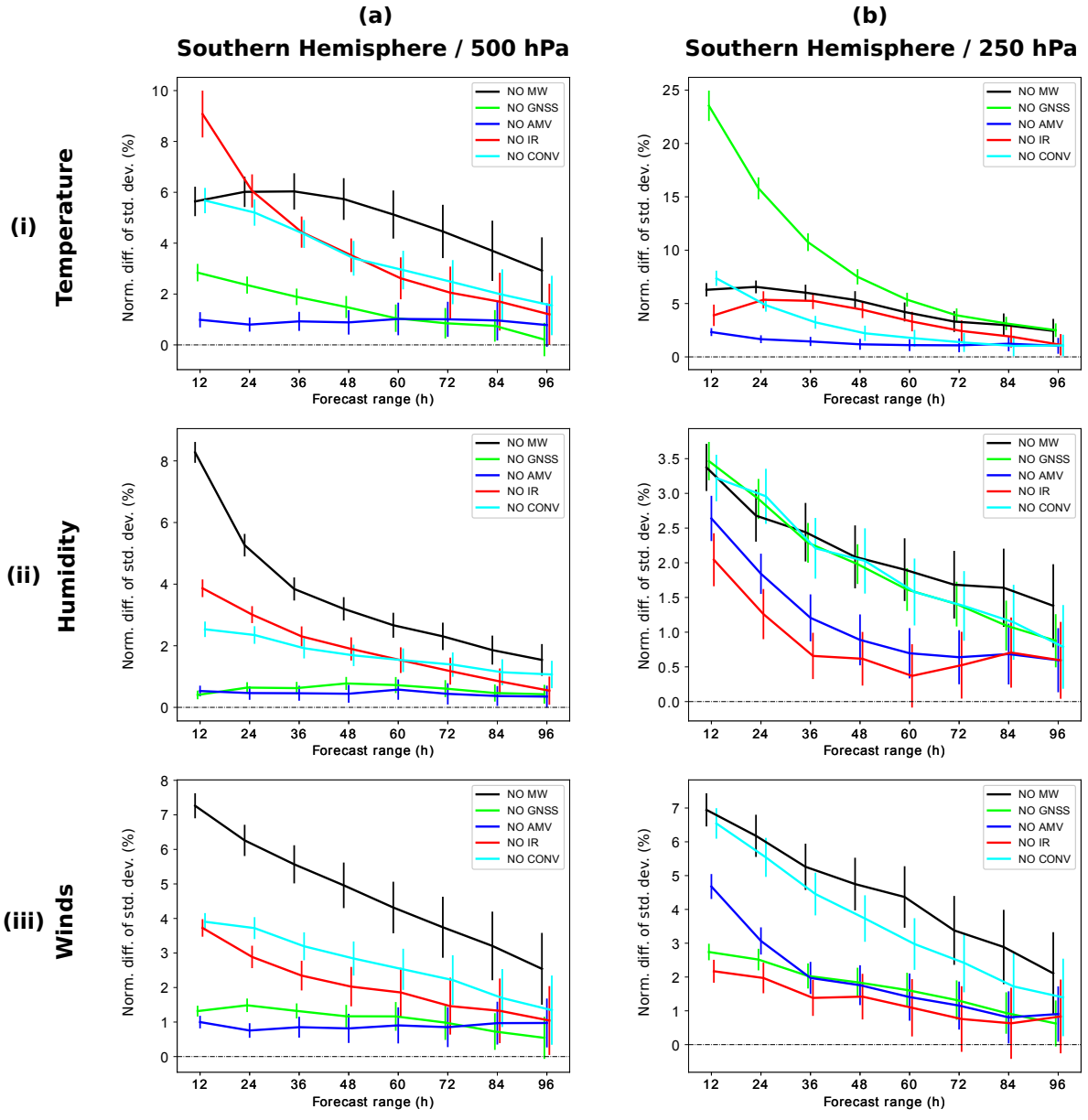


FIG. 4. Normalized difference in the standard deviation of the forecast error (against ECMWF analyses) in temperature (first row), relative humidity (second row) and wind vector (third row) versus the reference experiment **REF**, as a function of forecast range for five OSEs as listed in the legend. The left column corresponds to the Northern Hemisphere extra-tropics at 500 hPa, whereas the right column shows the tropics at 850 hPa. The period extends from October 2019 to March 2020 (6 months). The vertical bars indicate 99 % confidence intervals.

Forecast scores against ECMWF analyses expressed in terms of normalized standard deviation differences are compared for the first five denial experiments up to 96-h. Here, the focus is on the assessment of random error changes provided by the observations on forecasts.

Figure 4a shows them for temperature, relative humidity and winds at 500 hPa over the Northern Hemisphere. The most striking result is the very large degradation of the scores in the **NO CONV** experiment with values of above 14 % up to day-2 and of around 8 % on day-4 for wind and temperature. Microwave and infra-red radiances are the other major observing systems contributing to forecast skill scores with values of around 5 % during the first 24 hours. Their impact with respect to **NO CONV** is larger on humidity than on temperature and winds. The other observing systems GNSS-RO and AMVs have a much lower impact (of around 1 %) despite being significant up to the 48 h forecast range. These conclusions obtained at 500 hPa are very similar when examining other levels in the troposphere. The dominance of conventional observations on NWP forecast skill scores over the Northern Hemisphere has been identified in previous OSEs (e.g. (Bouttier and Kelly 2001; Radnóti et al. 2012). In the study of Bormann et al. (2019), the largest contribution of CONV data was noticed over mid-latitudes during winter, in agreement with our findings. Complementary experiments, to be shown in the next section, have been undertaken to examine more precisely the contribution of individual components of the conventional observing system (surface data, radiosoundings, aircraft reports).

In tropical regions, the lack of **IR** radiances significantly degrades the temperature at 850 hPa (Figure 4b) with values slightly above 6 % during the first 24 hours. The corresponding degradation induced by **NO MW** is smaller by a factor of two. There is even a slight improvement at short-ranges around 500 hPa (not shown). On the other hand, the largest negative impact on relative humidity at 850 hPa up to 60-h comes from the **NO MW** experiment. The importance of **AMVs** up to 36-h shows up clearly on humidity (likely from the horizontal transport) and on winds at 850 hPa (9 % after 12 hours). A similar behavior is noticed at 250 hPa regarding the impacts of **NO AMVs** on vector winds and relative humidity (not shown).



288 FIG. 5. Normalized difference in the standard deviation of the forecast error (against ECMWF analyses) in
289 temperature (first row), relative humidity (second row) and wind vector (third row) versus the reference experiment
290 **REF**, as a function of forecast range for five OSEs as listed in the legend. The left column corresponds to the
291 Southern Hemisphere extra-tropics at 500 hPa, whereas the right column shows the Southern Hemisphere extra-
292 tropics at 250 hPa. The period extends from October 2019 to March 2020 (6 months). The vertical bars indicate
293 99 % confidence intervals.

In the Southern Hemisphere at 500 hPa, the largest degradations are produced by the **NO MW** experiment on temperature, humidity and winds (Figure 5a). Conventional observations and infrared sounders contribute similarly but to a lesser extent to forecast skill score reduction, except in the short-range at 12-h for temperature with a larger loss of around 9 % from **NO IR**. One can see that at the 96-h forecast range the **NO MW** degradation on temperature remains above 3 % and is significant, whereas for the other experiments the normalized standard deviation is below 2 % with a reduced confidence level. Impacts which are almost negligible are noticed at that level for the **NO AMVs** experiment. The **NO GNSS** experiment leads to degraded scores on temperature of around 2 % during the first 36 hours, with a small corresponding impact on winds. At 250 hPa (Figure 5b), the impact of GNSS dominates temperature scores in the short-range up to 60 hours. The **NO MW** and **NO CONV** show similar behavior as at 500 hPa, whereas the degradation produced by **NO IR** is more reduced at shorter lead times. The degradation on temperature induced by **NO GNSS** impacts on relative humidity scores, similar to impacts seen in the **NO CONV** and **NO MW** experiments. The impact of **NO IR** on humidity at 250 hPa is smaller than that of **NO AMVs** which is likely induced by degraded advection forecasts. Winds scores at 250 hPa are mostly reduced by **NO MW** and **NO CONV** experiments, despite no direct measurements by microwave instruments and only a few in-situ wind measurements (aircraft, radiosoundings) in the Southern Hemisphere. Despite the few numbers in the Southern Hemisphere, radiosoundings provide invaluable vertical profile information that AMVs cannot bring. For example, Pourret et al. (2022) have shown the value of vertical wind profiles in data void regions from the Aeolus Doppler wind lidar despite its rather poor instrumental performances. The impact of **NO MW** on winds is caused by the strong coupling prescribed in the 4D background error covariance matrix at mid-latitudes , which allows the projection of temperature errors on wind errors from the accurate temperature profile retrievals observed by MW sounders. The lower impact of AMVs is caused by rather large observation errors specified in the 4D-Var system to account for uncertainities in the level height assignment. It is nonetheless comparable to that shown by Bormann et al. (2019). The remaining observing systems contribute to the error increase in a similar way (2 % in the short-range and no significance after day-3).

In summary, all observing systems provide useful information on NWP forecast skill scores. Those which have the largest generalized impact are CONV measurements and MW radiances

TABLE 2. Combined forecast skill scores (*IP18* index) over Europe averaged over a 6-month period (October 2019 - March 2020) for a baseline system (**REF**) and for various OSEs excluding conventional observations.

Experiment ID	**REF**	**NO CONV**	**NO RAOB**	**NO AIRCRAFT**	**NO SURF**
IP 18	6.51	-8.00	4.38	3.55	-1.63

despite representing only 17 % of the total observations which will be shown in a more quantitative way globally in Section 5. The IR radiances bring similar impacts as the MW but they are less pronounced. Bormann et al. (2019) argued that the actual MW constellation, having a large number of satellites with complementary orbits, leads to a more uniform coverage of the globe at each assimilation cycle than the IR constellation which is restricted to two main crossing equatorial times, as shown in Figure 2. GNSS-RO data dominate the temperature impact in the upper troposphere (and in the stratosphere) of the Southern Hemisphere and of the tropics in agreement with previous impacts studies such as those from Cucurull et al. (2007) and Bormann et al. (2019). Similarly, over the same regions, AMVs have large short-range impacts up to day-2 on vector wind forecasts at low and high levels. Such impacts project on the humidity field in the tropics through advective processes. Similarly, satellite radiances have an impact on extra-tropical wind forecasts; this effect is larger for **NO MW** in the Southern Hemisphere. These results show the ability of the 4D-Var system for extracting information from observations of one variable type and applying it to correct the background of a different variable type. Impacts on vector wind forecasts at low levels (up to 850 hPa) over all regions have also been observed with the **NO SCATT** experiment (not shown).

5. Complementary results

a. Contribution of conventional observations

Additional experiments have been undertaken by removing individual components of the conventional observing system:

- **NO AIRCRAFT**: aircraft reports (AIREP, ACARS, AMDAR) are excluded
- **NO RAOB**: radiosoundings, PILOT reports, wind profilers are excluded

- **NO SURF**: surface observations (SYNOP, BUOY, SHIP) in terms of geopotential, temperature, humidity, wind are excluded in the upper air analyses but are kept for the surface analyses in order to avoid any drift in land surface conditions in terms of soil temperatures and soil moisture contents.

In order to provide a quantitative analysis of these additional experiments, we use a specific NWP skill index defined at Météo-France to evaluate the model performances over Europe up to day-3. This NWP index called *IP18* considers three upper air parameters: 500 hPa geopotential, 850 hPa temperature and 200 hPa wind at two forecast ranges (48 and 72 h) issued from the 00 UTC analyses. For each parameter, the *RMSE* is computed against radiosoundings over Europe. It is then compared and normalized by its value in 2008 as $100 \times (RMSE_{2008} - RMSE)/RMSE_{2008}$. The global NWP skill index *IP18* is obtained by an arithmetic average of the six scores. The *IP18* values are displayed in Table 2 for the four OSEs. Positive values indicate improvements with respect to the NWP system in 2008. Removing all conventional observations has a large detrimental impact on forecast scores since the *IP18* index drops from 6.51 to a negative value of -8.00 (scores are significantly worse than the operational system in 2008 having a much reduced observing system and coarser NWP model resolution). The degradation is largest at 72 h on 500 hPa geopotential and on 250 hPa vector winds. When removing radiosounding data, the *IP18* is reduced to 4.38. This result reveals some resilience of the observing system since over mid-latitudes, aircraft reports and satellite radiances sensitive to temperature and humidity help to counteract the loss from radiosounding measurements. The loss of radiosoundings had a larger effect on short-range forecasts over Europe in the study of Bouttier and Kelly (2001) when satellite data and aircraft data where less numerous. Nowadays, the degradation over Europe when excluding aircraft data ($IP18$=3.55) is rather similar to the loss of radiosoundings showing the value of this observing system in regions well covered by commercial airlines. On the other hand, radiosounding data also provide information on humidity profiles which are not measured by aircraft over Europe and which is not accounted for in the *IP18* index. Finally, the largest degradation is induced the lack of surface observations, and more specifically on surface geopotential values (not shown) with a negative value of the *IP18* reaching -1.63. Indeed, surface pressure is known to be a key variable for mid-latitude weather forecasts, with no other observing system, apart from those in the CONV data category that can observe this quantity, to ensure resilience. Such observations (particularly

those reported by oceanic drifting buoys) always provide a large individual contribution in FSOI experiments (which will be demonstrated in a later section), despite their small numbers in the global observing system. This result is consistent with the fact that reanalysis systems with only surface observations (pressure and ocean winds) have been able to reconstruct realistic three-dimensional atmospheric fields when combined to a dynamical model within an advanced data assimilation system (Poli et al. 2016).

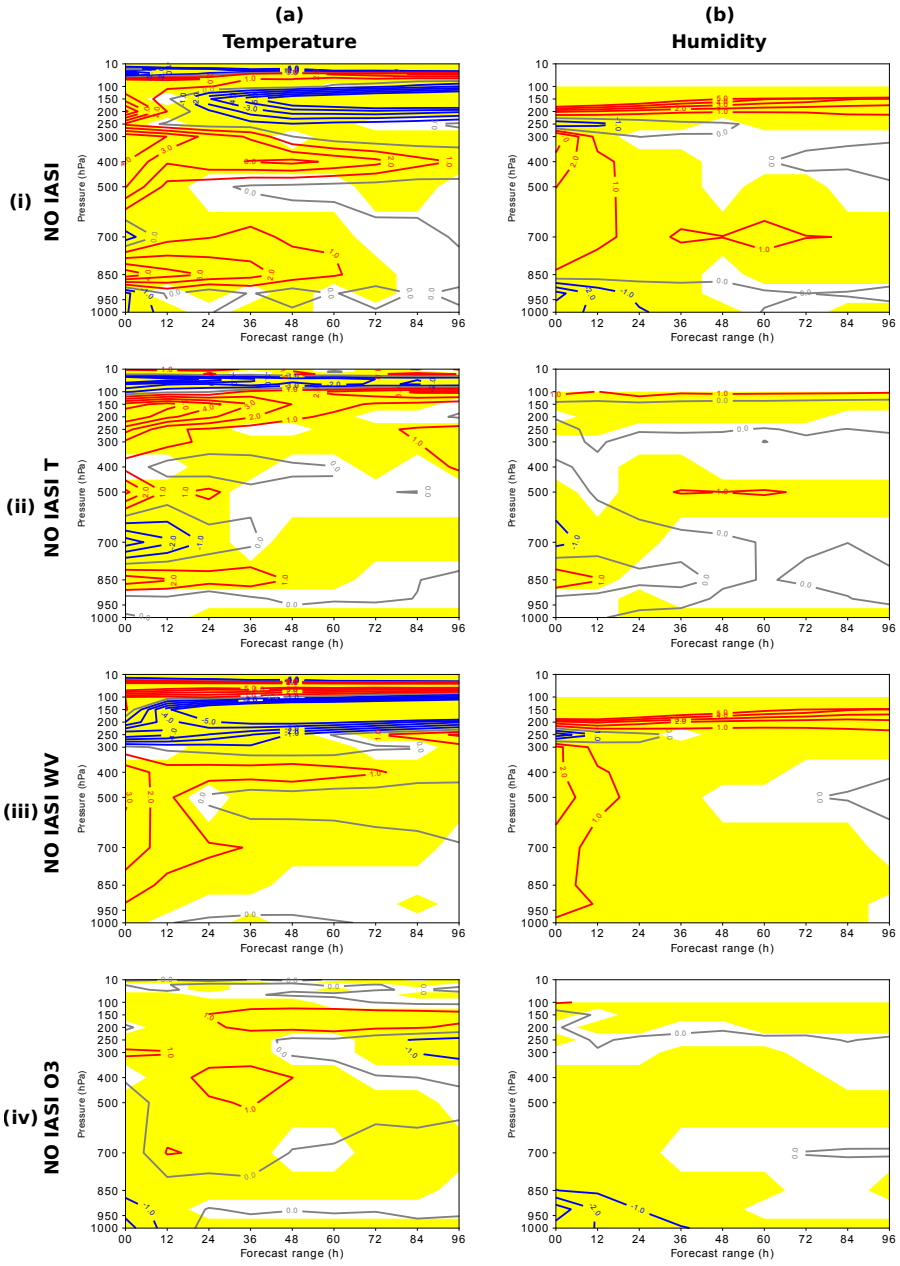


Fig. 6. Normalized RMSE values (in percent) for tropical temperature (left panels) and relative humidity (righ panels) against ECMWF analyses for denial experiments **NO IASI** (top row), **NO IASI T** (second row), **NO IASI WV** (third row) and **NO IASI O3** (bottom row) against a baseline system **REF** for forecast ranges up to 102 hours. Negative (red) values indicate a positive impact of the observing system (degradation of the forecast skill scores. Positive (blue) values indicate a negative impact of the observing system (improvement of the forecast skill scores.Yellow areas indicate where the differences are significant up to 99 % confidence. The period ranges from October 2019 to March 2020.

b. Contribution of infra-red radiances

Complementary experiments have been undertaken to examine more precisely the contribution of infra-red radiance denials:

- **NO IASI**: all IASI channels are excluded
- **NO IASI T**: all IASI channels sensitive to temperature (at most 97) are excluded
- **NO IASI WV**: all IASI channels sensitive to water vapor (at most 20) are excluded
- **NO IASI O3**: all IASI channels sensitive to ozone (at most 5) are excluded
- **NO GEORAD**: radiances from geostationary imagers are excluded

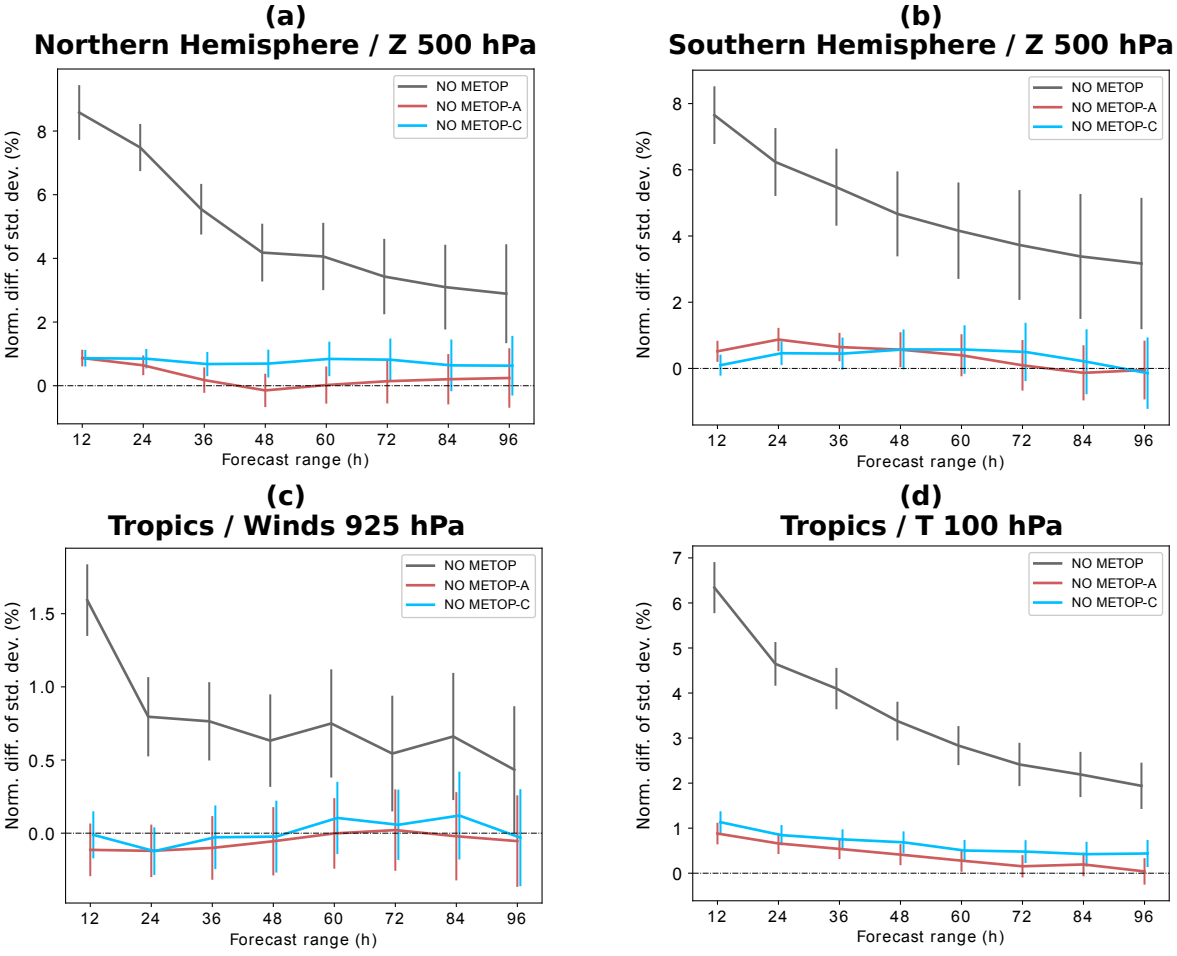
The results show that the **NO IR** signals described in the previous section are derived to a large extent from the three IASI instruments. The contribution of geostationary radiances is small but their availability at high temporal frequency (every 30 min in the 4D-Var) enables them to produce some wind forecast degradations when excluded (up to day-2 over mid-latitudes and up to day-4 in the upper tropical troposphere) (not shown). This impact is rather small since the instruments (imagers) have only got a reduced set of channels for assimilation (2 in the water vapor band and 3 in window regions). Despite being used at high temporal frequency, temporal correlation errors are not considered so far in the 4D-Var system which can lead to a sub-optimal usage. The possibility of extracting wind information from time-series of clear-sky radiances in a 4D-Var system is perhaps not totally consistent with AMVs which are more representative of cloudy regions. Additional studies to assess more precisely their complementarity should be undertaken.

Figure 6 displays the temperature and relative humidity forecast skill scores (Normalized *RMSE* values against the baseline system) over the tropics. The lack of IASI temperature channels leads to worse scores of temperature and relative humidity in the troposphere and upper stratosphere. Unexpected positive impacts on temperature are noticed however between 150 and 50 hPa in the extra-tropics (not shown) and around 700 hPa and 50 hPa in the tropics. The water vapor IASI channels impact the forecasts of mid-tropospheric relative humidity in the short-range and also in the upper troposphere at all ranges. On the other hand, despite short-range degradations below 500 hPa, temperature forecast improvements from the removal of IASI observations are observed in the upper troposphere (limited to the short-range over mid-latitudes but extending

over all forecast ranges in the tropics) around 200 hPa. A similar improvement from withholding IASI observations is noticed near 10 hPa. These mixed results regarding the use of IASI WV channels will require specific investigations, such as a revision of the current operational channel selection and the associated quality controls. By withdrawing ozone channels a slight positive and expected degradation takes place at high levels, however, wind, temperature and humidity are slightly improved in the lower troposphere. This is probably due to the use of a single climatological profile in the radiative transfer modeling, leading to a signal aliasing on other model variables. Coopmann et al. (2020) have recently obtained significant improvements on forecast scores of the ARPEGE model when using a more realistic ozone field in the radiative transfer model.

c. Resilience of observing systems

From the previous experiments the current observing system appears to be rather resilient to the loss of some components. It is remarkable that by withdrawing the IR radiances accounting for 80 % of the observations, the degradation of the forecasts is at the most 6 % in the short-range. On the other hand, conventional observations (7 % of all observations) can degrade up to 15 % Northern Hemispheric scores in the short-range, with a significant contribution from surface pressure observations.



435 FIG. 7. Normalized difference in the standard deviation of the forecast error (against ECMWF analyses) for the
436 extra-tropical (NH: left panel ; SH: right panel) geopotential at 500 hPa (first row), for the tropical wind vector
437 at 925 hPa (second row - left panel) and for the tropical temperature at 100 hPa (second row - right panel) versus
438 the reference experiment **REF**, as a function of forecast range for three OSEs where *Metop* satellites have been
439 excluded as shown in the legend. The period extends from October 2019 to March 2020 (6 months). The vertical
440 bars indicate 99 % confidence intervals.

Experiments have been undertaken where the three *Metop* satellites have been withdrawn (**NO METOP**) and where only one satellite is excluded (**NO METOP-A** and **NO-METOP-C**). In terms of extra-tropical scores (Figure 7ab), results from experiment **NO METOP** are very similar to those obtained with **NO IR** (due to the absence of 3 IASI instruments) but the degradation is lesser with respect to the **NO NW** denial experiment (only 6 microwave sounders being lost upon 18 instruments). In tropical regions, the absence of 3 scatterometers (among 4) and 3 GNSS-RO receivers (among 6) explains the score degradations noticed for the wind at 925 hPa and the temperature at 100 hPa respectively (Figure 7cd). These results are coherent with those obtained by McNally (2012) who examined the loss of polar orbiting satellites from Europe and USA on NWP forecast scores at ECMWF. On the other hand, excluding only one satellite leads to rather neutral results, scores being slightly worse with **NO METOP-C** which has more recent instruments (Figure 7). Such a result reveals that the end of life of *Metop-A* that took place in November 2021 has not been detrimental to the forecast skill scores of global operational NWP models.

d. Comparison with FSO impacts

The previous results can be presented in a synthetic manner by considering a global forecast error J based on the total energy norm expressed in $\mathrm{J.kg^{-1}.m^{-2}}$ and used classically for FSOI studies (Cardinali 2009). The use of a global energy norm allows the comparison of every meteorological variable at all model levels from the various OSEs with a single number.

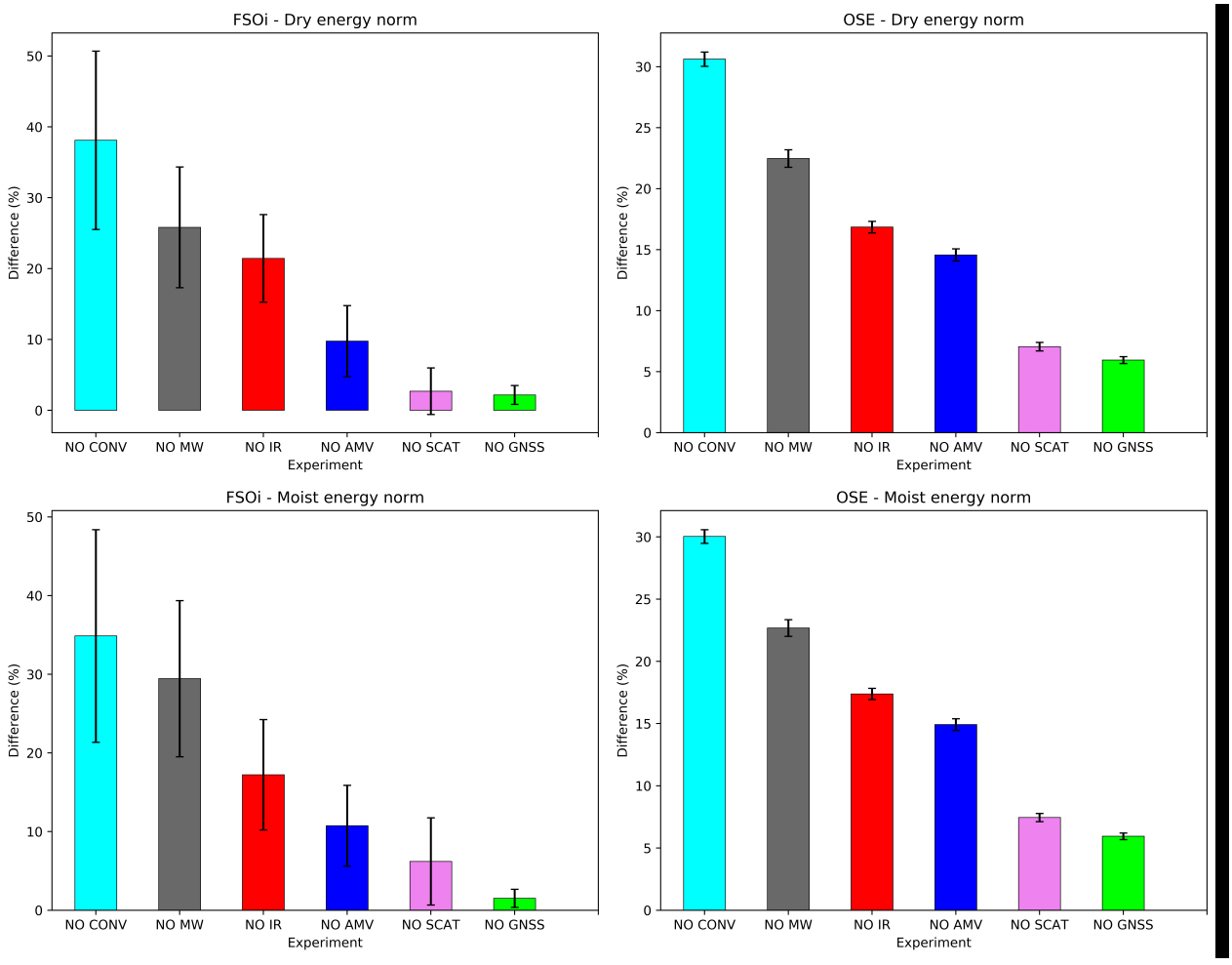
As previously performed by Gelaro and Zhu (2009), this direct measure of the forecast impact obtained in OSEs can be compared to that estimated by the FSOI technique using the adjoint of the forecast model and of the data assimilation system. Such comparison can help to gain confidence on FSOI results and one can examine whether or not they can be extended to forecast ranges beyond 24 hours. As pointed out by these authors and also more recently by Eyre (2021), when comparing the two methods differences should be expected due to their design in evaluating observation impacts. The FSOI measures the impact of observations with a background state containing information on all past observations. The OSEs measure cumulative effects of removing observations from both the background and the analysis.

The comparison is performed over a three-month period (January -March 2020) where the operational FSOI with the ARPEGE 4D-Var system had the same observing system as the OSEs (Table 1).

We have chosen for the OSEs the moist global energy norm used in operational FSOI at Météo-France and proposed by Ehrendorfer et al. (1999):

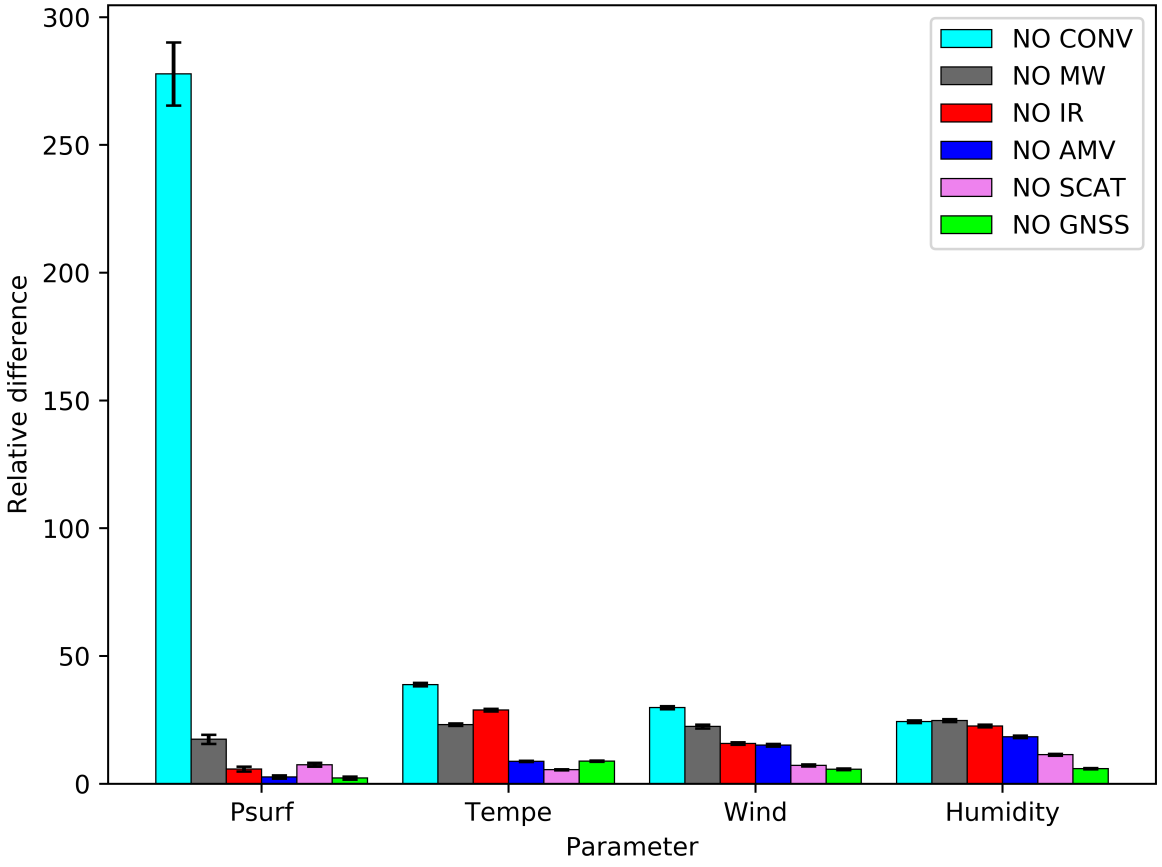
$$
\begin{aligned}
J &= \frac{R_d \mathrm{T_r}}{2g\mathrm{P_r}\Sigma} \iint (\mathrm{P}_{sf} - \mathrm{P}_{sa})^2 d\Sigma + \frac{1}{2}\frac{C_{pd}}{\mathrm{T_r}\Sigma} \iiint (\mathrm{T}_f - \mathrm{T}_a)^2 d\Sigma d\eta \\
&+ \frac{1}{2\Sigma} \iiint \left[(\mathrm{U}_f - \mathrm{U}_a)^2 + (\mathrm{V}_f - \mathrm{V}_a)^2 \right] d\Sigma d\eta + \frac{L_v^2}{2C_{pd}\mathrm{T_r}\Sigma} \iiint w_q (\mathrm{q}_f - \mathrm{q}_a)^2 d\Sigma d\eta \qquad (1)
\end{aligned}
$$

where R_d is the gas constant for dry air, C_{pd} is the specific heat of dry air at constant pressure, L_v is the latent heat of vaporization, $\mathrm{T_r}$ is a reference temperature (taken as 300 K), $\mathrm{P_r}$ is a reference pressure (taken as 1000 hPa). The empirical constant weight w_q is set to 0.3 for the moist energy norm and to zero for the dry energy norm. The integration extends on the full horizontal domain Σ and on the vertical using the hybrid vertical coordinate system η. For each prognostic variable (surface pressure $\mathrm{P_s}$, temperature T, wind components (U, V), specific humidity q), the subscript f corresponds to the forecast value at a given range. The subscript a corresponds to an analysis assumed to be a reasonable proxy of the true atmospheric state which is the one from the baseline experiment **REF** for the OSEs and a truncated low resolution version from the operational system ($\mathrm{T_L}224$) for the FSOI.



483 FIG. 8. Normalized adjoint (left panels) and OSE (right panels) based fractional impact of various observing
484 systems on the change in 24-h forecast error defined as a dry energy norm (top panels) and a moist energy
485 norm (bottom panels) over a 3-month period (January - March 2020).The vertical bars indicate 99 % confidence
486 intervals.

Figure 8 compares the forecast error increase at 24-h ($\Delta J = J_{EXP} - J_{REF}$) obtained from the set of six OSEs described in Section 3, together with that resulting from the operational Météo-France FSOI system ($\delta J = [\partial J/\partial y] \times \delta y$ where δy is the innovation vector). The ranking between these major observing systems is kept between OSEs and FSOI, the two most important being the conventional and microwave data followed by infra-red and AMVs. The fractional values compare well between OSEs and FSOI. The lowest contribution stems from GNSS-RO and SCATT associated with rather large confidence intervals for the FSOI. Indeed, they represent the smallest percentages in terms of observation number and affect rather specific regions of the atmosphere: ocean surfaces and upper troposphere/lower stratosphere. The impact of AMVs appears to be lower using the FSOI by a factor of two, since in the OSEs its contribution is close to that of the IR. Such a difference has also been noticed by Gelaro and Zhu (2009). The use of a moist energy norm has a non negligible impact on FSOI values for MW radiances since they contain many channels sensitive to water vapor. Such influence is also noticeable on SCATT (likely induced by a degradation of low level humidity advection). This effect is not present on AMVs because the contribution of the moist term in the upper troposphere (where the impact of these derived winds dominates) is very small since it is expressed in terms of specific humidity without vertical dependency (see Marquet et al. (2020) for a discussion on this point).



504 FIG. 9. Relative contributions to the 24-h forecast error on surface pressure (Psurf), temprature (Tempe),
505 horizontal wind components (Wind) and specific humidity (Humidity) expressed in terms of moist total energy
506 norm defined in Eq. (1) for six OSEs experiments against a baseline observing system experiment **REF** over a
507 three-month period (January-March 2020). The vertical bars indicate 99 % confidence intervals.

Examination of the individual components of the 24-h total forecast error expressed in terms of moist total energy norm (surface pressure, temperature, winds, specific humidity) for the 6 main OSEs (Figure 9) highlights the dominance of the **NO CONV** experiment on the surface pressure contribution 250 % increase. This large impact can be explained by the fact that these are relative differences. The absolute values for J_{REF} are respectively 3.1E2, 8.6E4, 9.5E5 and 1.0E5 for P_s, T, (U, V) and q, indicating that the pressure contribution changes are actually the smallest in absolute terms despite being the largest in relative terms. Moreover, all experiments are evaluated against the analyses of the **REF** experiment, which have small errors in the short-range. The **NO CONV** experiment also leads to the largest changes, but to a lesser extent, on other quantities. Microwave radiances have a contribution which is evenly spread among the four quantities, whereas the impact of infra-red radiances is larger on temperature and humidity. As previously noticed, the **NO AMVs** and **NO SCATT** experiments lead to a significant degradation of the humidity field. Finally, the **NO GNSS** experiment has its largest but relatively small impact on temperature (explained by the fact that the GNSS-RO measurements represent 0.5 % of the total observations).

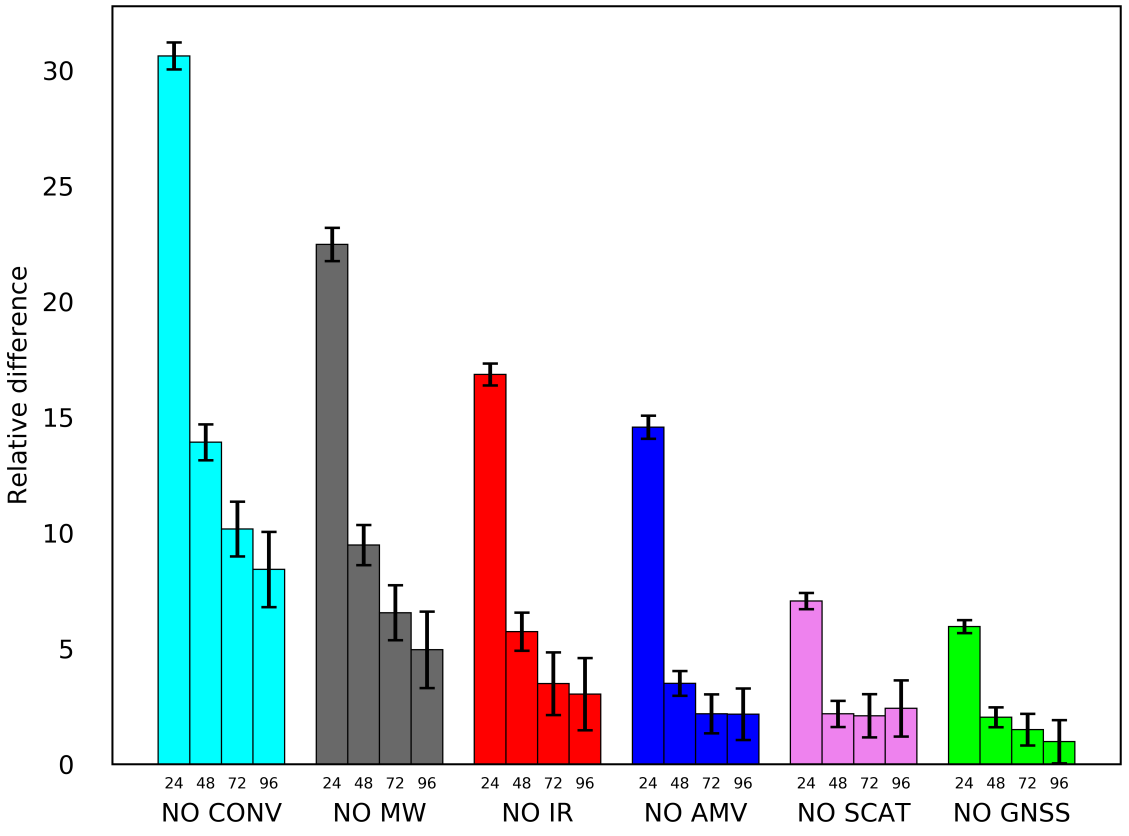


FIG. 10. Normalized OSE based fractional impact of various observing systems on the change in forecast errors (24-h, 48-h, 72-h, 96-h) defined as a moist energy norm over a 3-month period (January - March 2020). The vertical bars indicate 99 % confidence intervals.

When considering longer ranges (Figure 10), the impact of AMVs (and also SCATT and GNSS-RO but less pronounced) diminishes more rapidly than that of CONV and MW. It is interesting to see that the ranking of the three dominant observing systems identified in the short-range (24-h) by the FSOI is kept at longer ranges (96-h) in the OSEs.

6. Conclusions and recommendations

The global Météo-France NWP model ARPEGE and its 4D-Var data assimilation system have been used to undertake, in a configuration close to the current operational one, a series of OSEs to assess the impact of the global observing system on forecast skill scores. Experiments across a 6-month period have been performed at low horizontal resolution (factor of two with respect to the operational configuration) but with a comprehensive observing system (40 millions observations assimilated every day).

A number of key results consistent with previous studies have been obtained:

- The importance of conventional observations (despite their small fractional amount) in the Northern Hemisphere where they are the most numerous, but also over other regions. Surface pressure data are essential to avoid large forecast errors.

- Satellite radiances play a dominant role in tropical regions and in the Southern Hemisphere. They have a significant impact on mid-latitude winds (particularly MW radiances over the Southern Hemisphere). Microwave radiances also provide very useful information on atmospheric humidity and their impact remains significant at longer ranges (up to 96 h). Infra-red radiances also have a positive impact but which is less pronounced at longer ranges. Since they represent 80 % in terms of the number of observations, each individual radiance has a rather low information content. They are dominated by the 3 IASI sounders in terms of observation quantity and observation impact. Water vapor channels appear be detrimental at some locations, requiring further investigation.

- AMVs are particularly beneficial at low and high levels over the tropics and in the Southern Hemisphere mostly at short-ranges. Positive impacts have also been observed on the humidity field. The impact of SCATT winds is limited to low levels but is kept at longer forecast ranges.

- GNSS-RO bending angles improve the temperature in the high troposphere and low stratosphere outside the Northern Hemisphere. Their moderate impact comes from a reduced amount of receivers during the selected period (end of life of COSMIC-1 constellation and prior to the availability of COSMIC-2).

The comparison between FSOI and OSEs was made by examining a global forecast error based on the total energy norm at different forecast lead times. Results show a consistent ranking and relative contribution of the major observing systems (CONV, MW and IR). The impact of AMVs appears to be lower with FSOI diagnostics whereas the contribution of humidity sensitive observations (particularly for microwave radiances) is not straightforward in this context due to possible non-linearities of physical processes not properly handled by the adjoint method. The short-range impact highlighted by FSOI is kept at longer ranges for CONV, MW and SCATT observations.

All results have been presented in terms of mean forecast skill scores over large domains in order to draw robust conclusions. It could also be of interest to document, in future OSE studies, the observation impacts on high-impact weather quantities such as intense precipitation events or tropical cyclone tracks. This would however require conducting experiments over longer time periods to obtain reliable results.

These denial experiments confirm once again the important role played by conventional observations on the skill of NWP forecasts despite the growing availability and usage of satellite observations during the last two decades. Therefore, even though in-situ measurements can be expensive (e.g. radiosoundings in data void regions) they are vital to the quality of the Global Basic Observing Network (GBON) as defined by WMO. The recent WMO initiative SOFF (Systematic Observation Financing Facility) [11] to enhance surface and upper-air observations in developing countries by multipartner trust funds is particularly welcome (as shown clearly in Figure 2).

The impact of infra-red radiances despite being positive raises questions on how to best extract their information content since they represent by far the largest contribution in terms of percentage (80 %) but their withdrawal has less impact than **NO CONV** and **NO MW** experiments. Complementary satellite orbits could help to enhance their impact. An early morning orbit (Equatorial Crossing Time at 5:30 desc.) Chinese meteorological satellite FY-3E has been recently launched (July 2021) with on board an infra-red hyperspectral sounder HIRAS-2. Impact studies to be un-

[11] https://public.wmo.int/en/media/news/support-grows-systematic-observations-financing-facility

dertaken in the near future by assimilating radiances from this instrument should provide guidance on the interest of such an orbit to enhance the role of infra-red sounders for NWP. The difficulty of an optimal selection of radiances on instruments having more and more channels with correlated observation errors (e.g. the number of channels on IASI-NG to be launched by EUMETSAT in 2024 will be 16921) requires other methods to be explored in the NWP context. One can cite the decomposition of the full spectrum in Principal Components (PC) in order to assimilate the most informative PC scores (Matricardi and McNally 2014; Lu and Zhang 2019) or the assimilation of Level 2 retrieved profiles (Prates et al. 2016; Salonen and McNally 2020).

The large positive impact of microwave radiances on temperature, humidity and extra-tropical winds could be enhanced by their assimilation in cloudy/rainy areas within the ARPEGE 4D-Var as it is done nowadays in many operational NWP centres following the ECMWF initiative (Geer et al. 2017, 2018). A number of new satellite missions are planned in the coming decade in order to increase the temporal revisit of such measurements (constellations of nano or small satellites such as TROPICS (Blackwell et al. 2018) and AWS [12]). The exploitation of new frequencies of the microwave spectrum above 200 GHz sensitive to ice clouds (ICI on board EPS-SG) and below 19 GHz sensitive to precipitation and surface properties (radiometers from JAXA: AMSR-3 and ESA/Copernicus: CIMR [13]) should also contribute to the improvement of NWP forecast skills.

Radar scatterometers represent a unique observing system measuring ocean surface winds over wide areas (particularly in the tropics and Southern Hemisphere). They are only available however on few operational satellites, the longest time series being provided by the ASCAT instrument (C-band radar) on-board *Metop* (since 2006). A number of space agencies (ISRO, NSOAS, CNSA, NASA) have launched during the past decade scatterometers in Ku-band (a frequency that is more affected by precipitation) but with rather short durations (3 years in average) and/or issues with near-real time availability. In the context of the development of coupled atmosphere-ocean models, enhancing this observing capability in a sustainable fashion would be extremely valuable.

The importance of atmospheric wind measurements has also been highlighted in this study. Despite their small percentage (1.5 %) and their rather indirect estimation (cloud displacements), they are the most important remotely sensed observation contributing to the skill of wind forecasts in the short-range. Future satellite missions devoted to direct measurements of wind profiles through

[12]https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/Arctic_Weather_Satellite
[13]https://www.esa.int/ESA_Multimedia/Images/2020/11/CIMR

active sensors (lidars or radars) would likely benefit the NWP community. A more efficient direct extraction of wind information in data assimilation algorithms of coherent features (e.g. satellite radiances sensitive to water vapor or ozone) measured at high temporal frequency should be further studied, despite known limitations (Allen et al. 2013).

Finally, GNSS-RO data has a small, but positive impact which is likely because there were only a few of these observations (0.5 % of total counts) during our study period. The recent increase induced by additional receivers (6 from the equatorial COSMIC-2 constellation, KOMPSAT-5, GNOS/FY-3D, SEOSAR-PAZ) at Météo-France in July 2020 has significantly increased the impact of these data in the ARPEGE model (identified by specific OSEs and FSOI results). This has also been observed by other NWP centres. The interest in assimilating more data from GNSS receivers has been documented by ECMWF and the Met Office during the 2020 COVID-19 pandemic during which free access of data from the private company Spire was made possible. These results agree with the findings of Harnisch et al. (2013), which documented a possible saturation of GNSS-RO measurements for global NWP of 100,000 daily profiles that has not yet been reached. Making more data from GNSS-RO receivers available to the NWP community should be encouraged by space agencies because apart from their own value, such data are unbiased and thus allow a better usage of satellite radiances. They can also serve the operational space weather community by monitoring the activity of the ionosphere.

Acknowledgments. Dominique Puech (now retired from Météo-France) has been instrumental in developing the software packages that have been used to exploit the results of the experiments. The first version of this paper has been improved significantly thanks to the recommendations provided by the three reviewers.

Data availability statement. The numerical model and the data assimilation system are being developed at Météo-France, in collaboration with ECMWF and the European consortium for Limited Area Numerical Weather Prediction ACCORD. The code sources are not available under open source license. Datasets produced during the course of this study (ARPEGE analyses and forecasts) are too large to be publicly archived. All model and experiment data have been archived on the Météo-France mass storage system and can be obtained from the first author upon request.

References

Allen, D. R., K. W. Hoppel, G. E. Nedoluha, D. D. Kuhl, N. L. Baker, L. Xu, and T. E. Rosmond, 2013: Limitations of wind extraction from 4D-Var assimilation of ozone. *Atmospheric Chemistry and Physics*, **13 (6)**, 3501–3515, https://doi.org/10.5194/acp-13-3501-2013.

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

Blackwell, W. J., and Coauthors, 2018: An overview of the TROPICS NASA Earth Venture Mission. *Quart. J. Roy. Meteor. Soc.*, **144 (S1)**, 16–26, https://doi.org/https://doi.org/10.1002/qj.3290.

Bormann, N., H. Lawrence, and J. Farnan, 2019: Global observing system experiments in the ECMWF assimilation system. Tech. Rep. 839, ECMWF. https://doi.org/10.21957/sr184iyz, URL https://www.ecmwf.int/node/18859, 24 pp.

Bouttier, F., and G. Kelly, 2001: Observing-system experiments in the ECMWF 4D-Var data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **127**, 1469–1488.

Bouyssel, F., and Coauthors, 2022: *The 2020 global operational NWP data assimuilation system at Météo-France*, 645–664. Springer, Park S. K., Xu L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrological Applications (Vol. IV), https://doi.org/10.1007/978-3-030-77722-7_25.

Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135 (638)**, 239–250, https://doi.org/https://doi.org/10.1002/qj.366.

Coopmann, O., V. Guidard, N. Fourrié, and B. Josse, 2020: Use of variable ozone in a radiative transfer model for the global Météo-France 4D-Var system. *Quart. J. Roy. Meteor. Soc.*, **146 (733)**, 3729–3746, https://doi.org/https://doi.org/10.1002/qj.3869.

Courtier, P., C. Freydier, J.-F. Geleyn, F. Rabier, and M. Rochas, 1991: The Arpege project at Météo-France. *Seminar on Numerical Methods in Atmospheric Models, 9-13 September 1991*, ECMWF, Shinfield Park, Reading, Vol. II, 193–232, URL https://www.ecmwf.int/node/8798.

Courtier, P., and J.-F. Geleyn, 1988: A global numerical weather prediction model with variable resolution: Application to the shallow-water equations at Météo-France. *Quart. J. Roy. Meteor. Soc.*, **114 (483)**, 1321–1346.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120 (519)**, 1367–1387, https://doi.org/10.1002/qj.49712051912.

Cucurull, L., J. C. Derber, R. Treadon, and R. J. Purser, 2007: Assimilation of Global Positioning Systeme Radio Occultation observations into NCEP's global data assimilation system. *Monthly Weather Review*, **135**, 3174–3193.

Desroziers, G., P. Brousseau, and B. Chapnik, 2005: Use of randomization to diagnose the impact of observations on analyses and forecasts. *Quart. J. Roy. Meteor. Soc.*, **131 (611)**, 2821–2837, https://doi.org/10.1256/qj.04.151.

Duncan, D. L., N. Bormann, and E. Hólm, 2021: On the addition of microwave sounders and numerical weather prediction skill. *Quart. J. Roy. Meteor. Soc.*, **147**, 3703–3718, https://doi.org/10.1002/qj.4149.

Ehrendorfer, M., R. M. Errico, and K. D. Raeder, 1999: Singular-vector perturbation growth in a primitive equation model with moist physics. *J. Atmos. Sci.*, **56 (11)**, 1627 – 1648, https://doi.org/10.1175/1520-0469(1999)056<1627:SVPGIA>2.0.CO;2.

Eyre, J. R., 2016: Observation bias correction schemes in data assimilation systems: a theoretical study of some of their properties. *Quart. J. Roy. Meteor. Soc.*, **142 (699)**, 2284–2291, https://doi.org/https://doi.org/10.1002/qj.2819.

Eyre, J. R., 2021: Observation impact metrics in NWP: A theoretical study. Part I: Optimal systems. *Quart. J. Roy. Meteor. Soc.*, **147 (739)**, 3180–3200, https://doi.org/https://doi.org/10.1002/qj.4123.

Geer, A. J., and Coauthors, 2017: The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteor. Soc.*, **143 (709)**, 3189–3206, https://doi.org/https://doi.org/10.1002/qj.3172.

Geer, A. J., and Coauthors, 2018: All-sky satellite data assimilation at operational weather forecasting centres. *Quart. J. Roy. Meteor. Soc.*, **144 (713)**, 1191–1217, https://doi.org/https://doi.org/10.1002/qj.3202.

Gelaro, R., and Y. Zhu, 2009: Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus*, **61A**, 179–193.

Harnisch, F., S. B. Healy, P. Bauer, and S. J. English, 2013: Scaling of GNSS Radio Occultation impact with observation number using an Ensemble of Data Assimilations. *Mon. Wea. Rev.*, **141 (12)**, 4395 – 4413, https://doi.org/10.1175/MWR-D-13-00098.1.

Karbou, F., F. Rabier, and C. Prigent, 2014: The assimilation of observations from the advanced microwave sounding unit over sea ice in the french global numerical weather prediction system. *Monthly Weather Review*, **142 (1)**, 125 – 140, https://doi.org/10.1175/MWR-D-13-00025.1, URL https://journals.ametsoc.org/view/journals/mwre/142/1/mwr-d-13-00025.1.xml.

Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, **56 (3)**, 189–201, https://doi.org/https://doi.org/10.1111/j.1600-0870.2004.00056.x.

Lu, Y., and F. Zhang, 2019: Toward ensemble assimilation of hyperspectral satellite observations with data compression and dimension reduction using principal component analysis. *Mon. Wea. Rev.*, **147 (10)**, 3505 – 3518, https://doi.org/10.1175/MWR-D-18-0454.1.

Marquet, P., J.-F. Mahfouf, and D. Holdaway, 2020: Definition of the moist-air exergy norm: A comparison with existing "moist energy norms". *Mon. Wea. Rev.*, **148 (3)**, 907–928, https://doi.org/10.1175/MWR-D-19-0081.1.

Matricardi, M., and A. P. McNally, 2014: The direct assimilation of principal components of IASI spectra in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **140 (679)**, 573–582, https://doi.org/ https://doi.org/10.1002/qj.2156.

McNally, T., 2012: Observing System Experiments to assess the impact of possible future degradation of the global satellite observing network. Tech. Rep. 672, ECMWF. 20 pp.

Poli, P., and Coauthors, 2016: ERA-20C: An atmospheric reanalysis of the twentieth century. *J. Climate*, **29 (11)**, 4083 – 4097, https://doi.org/10.1175/JCLI-D-15-0556.1.

Pourret, V., M. Savli, J.-F. Mahfouf, D. Raspaud, A. Doerenbecher, H. Bénichou, and C. Payan, 2022: Operational assimilation of Aeolus winds in the Météo-France global NWP model ARPEGE. *Quart. J. Roy. Meteor. Soc.*, **148(747)**, 2652–2671, https://doi.org/ https://doi.org/10.1002/qj.4329.

Prates, C., S. Migliorini, L. Stewart, and J. Eyre, 2016: Assimilation of transformed retrievals obtained from clear-sky IASI measurements. *Quart. J. Roy. Meteor. Soc.*, **142 (697)**, 1697–1712, https://doi.org/https://doi.org/10.1002/qj.2764.

Radnóti, G., P. Bauer, A. Mc Nally, and A. Horányi, 2012: ECMWF study to quantify the interaction between terrestrial and space-based observing systems on numerical weather prediction skill. Tech. Rep. 679, ECMWF. 98 pp.

Salonen, K., and A. McNally, 2020: MTG-IRS Level 2 data assimilation into the ECMWF model. Tech. Rep. EUM/CO/15/4600001613/TA, EUMETSAT. 29 pp.

Simmons, A., and A. Hollingworth, 2001: Some aspects of the improvement in skill of numerical weather prediction. Tech. Rep. 342, ECMWF. 35 pp.

Tan, D. G. H., E. Andersson, M. Fisher, and L. Isaksen, 2007: Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM–Aeolus wind profiling mission. *Quart. J. Roy. Meteor. Soc.*, **133 (623)**, 381–390, https://doi.org/https://doi.org/10.1002/qj.43.