A global observing-system simulation experiment for the EPS–Sterna microwave constellation

Louis Rivoire, Robin Marty, Thomas Carrel-Billiard, Philippe Chambon, Nadia Fourrié, Olivier Audouin, Maud Martet, Camille Birman, Christophe Accadia, Jörg Ackermann

ORIGINAL ARTICLE

Journal Section

A global Observing System Simulation Experiment for the EPS-Sterna microwave constellation

Louis Rivoire[1] | Robin Marty[1] | Thomas Carrel-Billiard[1] | Philippe Chambon[1] | Nadia Fourrié[1] | Olivier Audouin[1] | Maud Martet[1] | Camille Birman[1] | Christophe Accadia[2] | Jörg Ackermann[2]

[1]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France

[2]EUMETSAT, Eumetsat Allee 1, 64295 Darmstadt, Germany

Correspondence
Philippe Chambon, CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France
Email: philippe.chambon@meteo.fr

Funding information
EUMETSAT, contract number EUM/COS/LET/21/12484

A constellation of microwave sounders named EPS-Sterna is under study at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), with the aim of complementing the backbone orbits of the global observing system in Low Earth Orbit. The satellites of this constellation would be similar to the Arctic Weather Satellite (AWS) which is being developed by the European Space Agency (ESA). The microwave sounder onboard AWS is equipped with temperature sounding channels around the 50 GHz oxygen absorption band, water vapour sounding channels around the 183 GHz and 325 GHz absorption bands as well as window channels at 89 and 165 GHz. An Observing System Simulation Experiment (OSSE) has been conducted at the Centre National de Recherches Météorologiques (CNRM) to evaluate the impact of this constellation on Numerical Weather Prediction (NWP) at the global scale with the ARPEGE (Action de Recherche Petite Echelle Grande Echelle, Bouyssel et al. (2022)) model. Two periods ranging from August to October 2021 and December 2021 to February 2022 have been chosen to compute the nature run and to run 4D-Var data assimilation experiments. As validation of the OSSE framework, the impact of a Metop-B denial experiment in the OSSE is compared to the impact of a Metop-B denial with real observations. This comparison shows that the Metop-B denial impacts are very similar in the OSSE and with real observations, with the OSSE slightly overestimating the impact. Then, the impact of various scenarios for the EPS-Sterna constellation are assessed by computing forecast errors, fractions skill scores and moist global energy norms, and comparing to the results of a baseline experiment without the EPS-Sterna constellation. Significant and positive improvements of the forecasts are found up to 96h, for every variable tested, with an impact increasing with the number of satellites.

KEYWORDS

Data assimilation, Observing System Simulation Experiment, EPS-Sterna, Arctic Weather Satellite

1 | INTRODUCTION

Microwave sounding instruments, usually flying on high-performance and large platforms, have proven to be a major source of information for Numerical Weather Prediction (NWP) by providing temperature and humidity sounding information in all weather conditions. With the improved usage of these data, and in particular the advent of all-sky assimilation (Geer et al., 2017), their impact on forecast skill drastically improved (Duncan et al., 2021). Unfortunately, the number of available microwave sounders will likely decay in the next few years, with several ageing platforms being at the very end of their life time. Yet, with the advent of the miniaturisation of the sensors, small satellites are envisioned to complement the CGMS (Coordination Group for Meteorological Satellites) 3-orbit baseline and to improve the temporal resolution of the observations. In particular, in 2024 ESA plans to launch the AWS satellite (cf the ESA website ESA, 2023a). This small satellite will carry a 19-channel microwave radiometer, providing temperature (near 50 Ghz) and humidity (near 183 Ghz) soundings of the atmosphere. Additionally, the AWS instrument will carry four sounding channels at 325 GHz, not flown before on a space-borne instrument. EUMETSAT, together with ESA, is currently studying the added value of launching a constellation of small satellites carrying microwave sounders, called the EPS-Sterna constellation, with satellites very similar to AWS.

In order to support decision making by space agencies, evaluations of the potential benefits of envisioned future satellites is crucial. Different methodologies exist to measure these benefits, such as Ensemble Data Assimilation (EDA) or Observing System Simulation Experiments (OSSE) (Bormann et al., 2023). In this article, we assess the impact of the EPS-Sterna constellation using the OSSE methodology, building on Météo-France's past experience conducting OSSEs for other observing systems (Guedj et al., 2014; Duruisseau et al., 2017; Coopmann et al., 2023). An OSSE mimics a real NWP system with simulated data by providing forecasts with realistic errors compared to a known truth. Existing observations as well as new satellite instruments are simulated, thus deriving the impact of such new measurements on the quality of the forecasts and eventually comparing the utilities of competing designs. Such a methodology has already been developed in different institutes (Errico et al., 2013) and has already given insightful evaluations of the impacts of future observing systems, e.g. Stoffelen et al. (2006); Masutani et al. (2010); Privé et al. (2022).

This article presents an evaluation of the potential impact of the EPS-Sterna constellation on NWP, using an OSSE generated with the global model ARPEGE. Various scenarios are studied, with the number of satellites ranging from three to eight. First of all, an overview of the OSSE framework, its construction and its calibration is given in section 2. Then, the experimental set-up and verification methods are presented in section 3. Finally, the validation of the OSSE and the results of the impact study are given in section 4. Conclusions are then given in section 5.

2 | OBSERVING SYSTEM SIMULATION EXPERIMENT FRAMEWORK

2.1 | General framework

An OSSE is a replica of a real NWP system, in an entirely simulated environment. Thus, it has the capability to assimilate existing observation types but also envisioned ones. It can also compare the realistic forecasts to the simulated atmosphere, which is perfectly known. It requires several ingredients:

- A nature run, considered as true state of the atmosphere for observation simulation and forecast verification. It consists of a long, uninterrupted forecast and produces a realistic evolution of the atmosphere.
- Simulated observations, generated using the nature run, to which an appropriate noise is added to simulate mea-

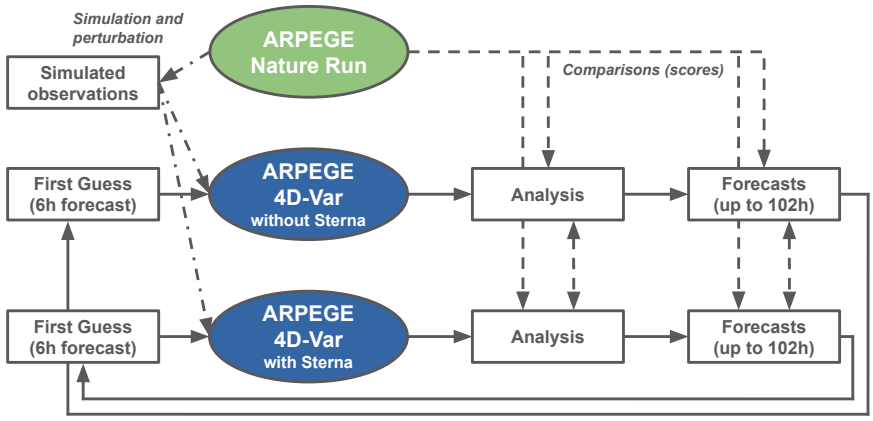


FIGURE 1 Description of the framework of the OSSE, with and without the EPS-Sterna constellation.

surement errors.

- A NWP model, used to compute the best estimates of the atmospheric state given the simulated observations and to produce weather forecasts.

A simplified description of the OSSE framework is provided in figure 1. The forecasts generated by the NWP model for a given assimilated observing system are compared with the nature run, which is used as reference for verification, to obtain forecast errors. Then these errors are compared among the experiments with different observing systems.

2.2 | The nature run

The nature run consists of a long and uninterrupted forecast. It is used to provide a realistic evolution of the atmosphere, which serves as a reality for our study. We consider that this simulated atmosphere is the true state of the atmosphere and we use it to simulate the observations (first row in figure 1), to initialise the assimilation (not shown on figure 1) and to verify the forecasts produced by the NWP model (second and third rows in figure 1).

The nature run is generated with ARPEGE cycle 46t1 (Bouyssel et al., 2022), which is the global NWP model operationally used at Météo-France since June 2022. One specificity of ARPEGE is that it is characterised by a tilted stretched grid. Using the spectral truncation TL1798 and a stretch factor of 2.2, which is the current operational

setup, leads to a horizontal resolution of about 5 km over Europe and about 24 km at the antipodes. The physical parametrizations used within the nature run are further described in Bouyssel et al. (2022); one main difference with the parametrizations presented in this paper is the use of the Tiedtke convection scheme (Tiedtke, 1989) which has been implemented in the Integrated Forecasting System (IFS) at ECMWF for many years and is now also used in the operational version of ARPEGE. A daily sea surface temperature (SST) forcing has been set up using OSTIA analyses (Operational Sea Surface Temperature and Sea Ice Analysis Stark et al. (2007)) in order to ensure a realistic SST evolution over the period.

Two four-month datasets are generated with outputs every 30 minutes, each of them having a 1-month spin-up period: the first from July to October 2021 (July is the spin-up month) and the second from November 2021 to February 2022 (November being the spin-up month).

Several validation procedures were performed on the nature run. In particular, one verification was to check that the mean fields were not strongly deviating from the operational forecasts after several months of independent run time. Mean global differences between the nature run atmospheric fields (temperature and humidity) and the operational ARPEGE analyses have therefore been performed for the two periods. In the case of temperature at 900 hPa, the mean global differences do not exceed 0.5 K (0.41 K after 2 months for the summer period and -0.11 K after 2 months for the winter period). Note that ARPEGE has already been used and validated for seasonal predictions (Batté and Déqué, 2016) and climate predictions (Roehrig et al., 2020), which includes a validation of its representation of clouds and precipitation against observations.

2.3 | The data assimilation system

In this study, two flavours of the same data assimilation (DA) framework are constructed: one with the ability to assimilate simulated observations, the other using real observations for validation purposes. The first is presented in the second and third rows of figure 1. The second uses real observations, and is created to conduct an Observing System Experiment (OSE), in order to validate the OSSE framework, as presented in section 4.1. The common aspects between the two flavours are presented as follows:

- A 4D-Var system using 6 h cycling and 30 min time slots. The analysis performed every six hours uses a relatively long cut-off (the cut-off is the time at which the analysis computation begins), of +3 h after the end of the assimilation window, long enough to collect the vast majority of available measurements before starting the computation.
- Everyday at 00:00 UTC, one long forecast up to +102 h is performed. This long forecast is based on an analysis which uses a relatively short cut-off (02:15 UTC) as ARPEGE in operations, thus collecting less measurements than the analysis performed with a long cut-off.
- The forecasts are run at a lower resolution than the nature run at spectral truncation TL798 and stretch factor 2.2, which leads to a resolution of about 10 km over Europe and about 61 km at the antipodes.
- The radiative transfer code RTTOV-SCATT v12 (Saunders et al., 2018) is used as the forward operator for radiance assimilation.
- The convection scheme used for forecasts as well as trajectories in the assimilation is the so-called Bougeault convection scheme (Bougeault, 1985), and was the scheme operationally used at Météo-France before switching to the Tiedtke convection scheme.

However, several differences are introduced between the OSSE and OSE in order to make the OSSE numerically

cost effective:

- Only the atmospheric data assimilation system is run for the OSSEs and not the surface data assimilation. In the OSSE experiments, the surface fields are down-scaled from the nature run.
- background errors are used for all OSSE experiments, compared to the flow-dependent background errors which are used in operations (Berre and Desroziers, 2010; Berre et al., 2015).
- No bias correction scheme is used in the OSSE framework. The bias of first-guess departure distributions was indeed checked for several observation types and sensor kinds and was considered negligible.

One pitfall that needs to be avoided when building an OSSE framework is the so-called “identical twin” problem (presented in Stoffelen et al. (2006) or Masutani et al. (2010)). If the exact same model is used for the nature run and data assimilation and forecasting system then the simulated observations effectively match the forecasts, which spuriously under-represents model error leading to erroneous impact assessments. In order to avoid this problem, several differences were designed between the nature run and the data assimilation system. This type of framework, consisting of the same model ARPEGE but with various differences, is called a "fraternal twin" OSSE. The nature run is configured at the higher TL1798 resolution with Tiedke convection compared to the data assimilation and forecasting system at the lower TL798 resolution with Bougeault convection, enabling an appropriately larger model error growth in the forecasting system. The forecast error growth obtained with this setup will be compared to a reference error growth in section 4.1.

2.4 | Simulation and calibration of a synthetic observing system

Numerous observations are simulated in the OSSE in order to represent a real distribution of observations as being used by the real NWP model. A list of the observing systems considered in the control experiment is given in table 1. Of all the observations being operationally assimilated in ARPEGE in 2021, several instruments have been removed from this baseline in order to mimic the observing system foreseen for 2030, when the EPS-Sterna constellation could be at the start of its deployment.

Among the discarded instruments, the Doppler Wind Lidar AEOLUS and some microwave sounding instruments have been removed and are thus not simulated. A follow-on mission to the Doppler Wind Lidar is currently being considered, similarly to the EPS-Sterna program, therefore AEOLUS has been removed in order to focus on discerning the actual contributions coming from EPS-Sterna. Similarly, at the beginning of this study, no satellite was planned for 2030 at the local time of ascending node for the following instruments: MWHS2 onboard FY-3C, SSMIS onboard DMSP-F18, AMSU-A and MHS onboard NOAA-19. Thus, these instruments have not been included in the baseline observing system in order to measure the impact of the EPS-Sterna constellation. As can be seen in table 1, the baseline includes the equivalent of 7 microwave sounding platforms with both temperature and humidity channels: ATMS (temperature and humidity) on board SNPP and NOAA-20, the combination of AMSU-A (temperature) on board NOAA-15 and SSMIS (humidity) on board DMSP-F17, AMSU-A (temperature) and MHS (humidity) on board NOAA-18, Metop-B anc Metop-C, and MWHS-2 (temperature and humidity) on board FY-3D.

In the version of ARPEGE used in this study, MHS, GMI, AMSR2 and MWHS-2 are assimilated in all-sky conditions (as presented in table 1), following the ECMWF methodology (Geer et al., 2017). The microwave radiometers assimilated use the default hydrometeor optical properties from RTTOV-SCATT v13.0 (Geer et al., 2021).

To simulate the observations, we follow this procedure: (i) we use the geographical and temporal location of real observations for the geographical location of simulated observations, in order to ensure a realistic sampling of the

Measurement type	Instrument
Surface	Surface stations, ships and buoys
	Wind profilers
Altitude	Radiosondes
	Aircraft measurements
Satellite	Atmospheric Motion Vectors (AMVs) from Meteosat, GOES, Himawari
	Clear-sky radiances from Meteosat, GOES, Himawari
	Clear-sky microwave sounding data from ATMS (onboard NOAA-20 and SNPP), AMSU-A (onboard NOAA-15, NOAA-18, Metop-B and Metop-C), SSMIS (onboard DMSP-F17)
	All-sky microwave sounding data from MHS (onboard NOAA-18, Metop-B and Metop-C), GMI (onboard GPM), AMSR2 (onboard GCOM-W), MWHS-2 (onboard FY-3D)
	Hyperspectral infrared data from IASI (onboard Metop-B and Metop-C), CrIS (onboard NOAA-20)
	Scatterometer winds
	GNSS radio occultation data

TABLE 1 Description of the various observing systems simulated and assimilated in the control experiment.

atmosphere, except for the Atmospheric Motion Vectors (AMVs) for which the geographical locations do not match the clouds in the OSSE but the real clouds as in operationally used data, (ii) we apply the observation operators from the ARPEGE operational code to the nature run, (iii) we add Gaussian noise to the simulated observations in order to account for instrumental errors. The noise is assumed to be uncorrelated for most observations, except for the hyperspectral infrared instruments for which correlated noise is considered.

Using the observation location of real observations ensures a realistic geographical distribution of the data without the use of an orbital simulator for each satellite, or the use of a flight simulator for aircraft data, etc. It also only takes into account the observations which have been received in time for being assimilated within the operational ARPEGE system. It ensures that all observing systems have their time of dissemination taken into account, with the time of dissemination being the latency between measurement and availability in the model.

A calibration of the noise added to the simulated observations is performed in order to have a weight in the assimilation similar to that of real observations. The calibration procedure applied is similar to the one adopted for the previous OSSEs conducted at Météo-France (Guedj et al., 2014; Duruisseau et al., 2017; Coopmann et al., 2023). The calibration is performed by tuning the magnitude of perturbations added to the simulated observation. This is an iterative process, which aims at having no statistically significant differences at the 99% confidence level between the first-guess departures statistics of real and simulated observations. The calibration ensures that each observing system will have impacts in the OSSE similar to the impacts of a real NWP system. Indeed, if the simulated observations are not perturbed enough, they will then be characterized by too low first-guess departures statistics. This means that the simulated observations will be too accurate, the analysis and forecasts will be overly constrained and thus be much more difficult to improve. On the other hand, if the perturbations assigned to simulated observations are too high (thus giving large first-guess departure statistics), the observations will not be accurate enough. In this case the analysis and forecasts will not be constrained enough, thus adding a new observing system will have an impact larger than what it would have in reality. Note that ideally, the instrumental noise of each observing system should be used as noise for the simulated errors, however they are not necessarily accessible for all the observation kinds.

The calibration process is iterative as mentioned above and follows the procedure presented hereafter: (i) we simulate an initial set of observations, using specific perturbations for each observation, (ii) we run a 15-day data

assimilation experiment, using the initial set of simulated observations and the assimilation framework presented in section 2.3, (iii) we compute the standard deviation of first-guess departures over the 15 days and we compare them with real first-guess departures from real observations, (iv) we derive new perturbations and a new set of simulated observations.

This process has been repeated seven times. As an example, for MHS channel 5 in the OSSE, the noise applied after the seven iterations is characterised by a standard deviation of 0.23 K; this number can be compared to the Noise Equivalent Delta Temperature (NeΔT) of MHS channel 5 onboard Metop-B and C around 0.15 K when a superobbing is taken into account. In this particular case, the noise derived from the calibration is therefore higher than the NeΔT of real observations. This comparison has been performed on a number of cases (not shown) and the noises derived from the calibration are sometimes very close to the instrumental noise, and sometimes lower or higher than the instrumental noise but always of the same order of magnitude. These differences between real instrumental noise and simulated instrumental noise give evidence that the differences between the nature run and the OSSE do not fully match the differences between a real NWP system and reality as will be evaluated in section 4.1.

With this calibration, the first-guess departure statistics presented in figure 2 are obtained. In this figure, positive values indicate larger first-guess departure standard deviations for the simulated data than for real data. Most of the blue error bars (after calibration) presented on this figure cross the zero line, which indicates that the calibration procedure has reached the point that there are no more statistically significant differences between first-guess departure standard deviations, for conventional and satellite observations. Note that in some cases, first-guess departure standard deviations of simulated data are still slightly above or below those of real observations. This means that the noise amplitudes may be further tuned by other rounds of calibration. However, one of the most important aspects that the calibration should result in is to obtain realistic impacts of observations, and in particular at forecast ranges longer than 6 h.

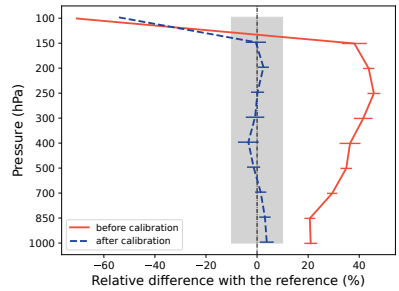
Two aspects will therefore be checked: the first one is the forecast error growth of the control experiment and the second one is the impact of a denial of one observing system. This validation will be performed by comparing the results from the OSSE, which uses simulated observations, and the OSE, which uses real observations. These two aspects will be described in section 4.1.

2.5 | Simulation of the EPS-Sterna constellation

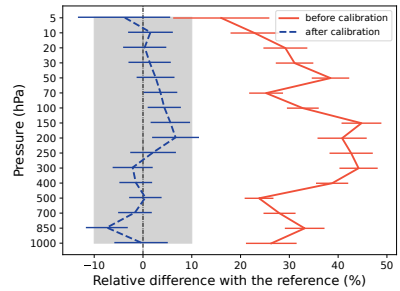
The EPS-Sterna constellation consists of small satellites orbiting at 595 km of altitude, each of them carrying an instrument with microwave sounding capability further described below. Five different scenarios for the EPS-Sterna constellation have been considered. They are named using the following syntax: OPn-mSAT, with “n” being the number of orbital planes on which the satellites are distributed and “m” being the total number of satellites in the scenario. A detailed description of the scenarios explored can be found in table 2. Note that in the selected baseline, the remaining microwave sounders are characterised by planes with local times at descending node (LTDNs) of 01:30 (FY-3D, NOAA-20, SNPP, so called afternoon orbits), 06:30 and 07:00 (DMSP-F17, NOAA-15, so called early morning orbits) and 09:30 (NOAA-18, Metop-B, Metop-C, so called mid morning orbits) (CGMS, 2022). The LTDNs of EPS-Sterna planes (03:30, 05:30, 07:30 and 11:30) therefore complement the baseline well.

Unlike for the other simulated observations, the geographical locations and times of observation have been derived from an orbital simulator for each scenario.

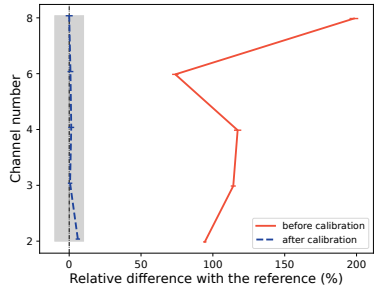
With 19 channels centred around 50 GHz, 183 GHz and 325 GHz (plus complementary window channels), the instrument on board these satellites has both a temperature and a humidity sounding capability, combining the strength of AMSU-A and MHS sensors onboard Metop. Details on the channels of the instrument of EPS-Sterna, including



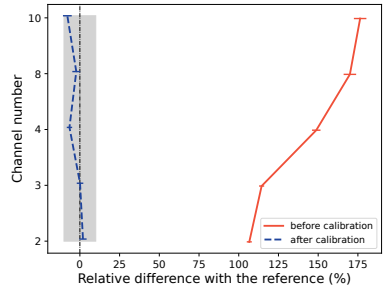
(a) aircraft reports, zonal wind



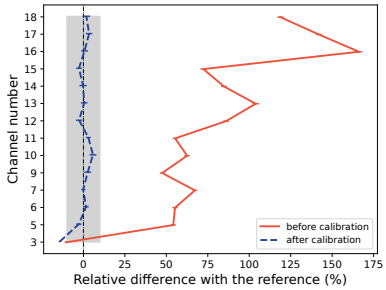
(b) radiosondes, temperature



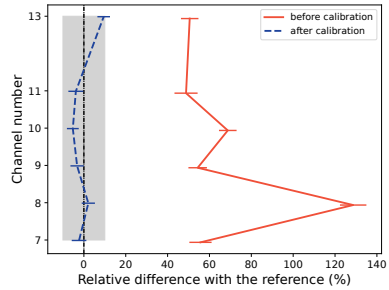
(c) SEVIRI onboard MSG



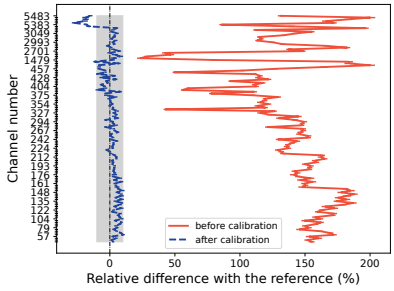
(d) AHI onboard Himawari



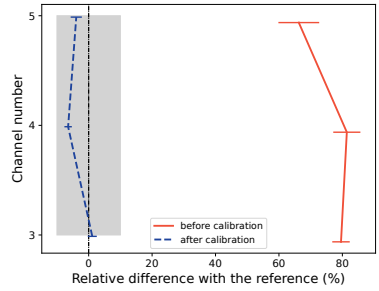
(e) SSMIS onboard DMSP-F17



(f) AMSR2 onboard GCOM-W, clear-sky data only



(g) IASI onboard Metop-B



(h) MHS onboard Metop-B and Metop-C, clear-sky data only

FIGURE 2 Relative first-guess departure standard deviations computed over the globe during 18 days from 14/08/2021 to 31/08/2021, for various observing systems after quality control. The zero line corresponds to first-guess departure standard deviation of real observations. Error bars indicate statistical significant results at the 99% confidence level; in the grey shaded areas, points are at a distance from the reference lower than 10%.

Scenario name	Nb of satellites	Planes	Repartition of the satellites over planes
OP3-3SAT	3	P1, P2 and P3	1 satellite per plane
OP2-4SAT	4	P1 and P2	2 satellites per plane
OP3-4SAT	4	P1, P2 and P3	2 satellites on P1, 1 satellite on P2, 1 satellite on P3
OP3-6SAT	6	P1, P2 and P3	2 satellites per plane
OP4-8SAT	8	P1, P2, P3 and P4	2 satellites per plane

TABLE 2 Description of the various scenarios used for the EPS-Sterna constellation in terms of orbits. The Local Time of Descending Node of the planes are the following: 03:30 LTDN for plane 1 (P1), 11:30 LTDN for plane 2 (P2), 07:30 LTDN for plane 3 (P3) and 05:30 LTDN for plane 4 (P4). When two satellites are on the same plane, the relative phasing between them is 180 degrees. The relative phasing between satellites on different planes is also optimised : e.g. the relative phasing is shifted by 90 degrees between the satellites on P1 and P2 in the OP2-4SAT scenario.

the individual NeΔT values, can be found in table 3 and more characteristics can be found on the ESA website (ESA, 2023b). The EPS-Sterna observations are simulated using the same process as described in section 2.4. The simulated observations are perturbed using the listed NeΔT divided by a factor of three, to account for a 3x3 superobbing of the observations (raw NeΔT corresponding to a sampling time of 0.0025 s). This superobbing has been applied to EPS-Sterna observations the same way it is applied to the observations of other microwave instruments assimilated within ARPEGE. Compared to the other observing systems, EPS-Sterna observations are therefore the only one to have their real instrumental noise taken into account, as there is no point of comparison with real observations. This adds an uncertainty onto the evaluation of EPS-Sterna impact within the OSSE framework.

Channel 1 (50.3 GHz), channel 2 (52.8 GHz), channel 3 (53.246 GHz), channel 9 (89 GHz) and channel 10 (165.5 GHz) have not been actively assimilated in this study as a first approach to focus on the impact of the upper tropospheric channels, insensitive to the surface. Channels 16 to 19 (325 GHz) have also been discarded from the assimilation and will be the subject of a future study dedicated to sub-millimetric frequencies. In terms of quality control, a horizontal thinning of 100 km is applied to the simulated orbits to avoid horizontal observation error correlation, and observations with a brightness temperature lower than 243 K are not assimilated in order to avoid cold surfaces. Observations over lakes and coastal areas are also avoided. Low-peaking channels are not assimilated over high mountains, depending on their weighting functions. Figure 3 shows OP3-6SAT observations for channel 11 (176.311 GHz) during a 6 h cycle, and which passed the various quality controls. Channels 4 to 8 and 11 to 15 are assimilated using the all-sky route developed at ECMWF. In order to do so, observation error models (Geer and Bauer, 2011) have been tuned for each channel of the instrument, choosing as initial implementation the same scattering index (difference between channels 9 and 10) for all channels and all surface type. Note that this cloud predictor has been well used for humidity channels but is likely sub-optimal for temperature channels (Duncan et al., 2022); this could be improved in a future assessment of EPS-Sterna as it will be mentioned in section 5. Various observation error model parameters have been tested (not shown) and the set of observation errors providing the best compromise have been retained for this study. The clear-sky observation errors used over oceans and land are presented in table 3. Note that the clear-sky errors for humidity channels are relatively high in order to minimize a degradation in the short range in the Tropics, which will be further described below.

For the production cycle centered at 00:00 UTC each day (which is used to compute the +102 h forecasts), only the Sterna observations between 21:00 UTC the day before and 00:45 UTC are used, in order to replicate the ARPEGE operational NWP context. Indeed at Météo-France, the ARPEGE operational run has to be launched at 02:15 local

Channel nb	Frequency (GHz)	Used NeΔT (K)	Clear-sky obs. error on ocean (K)	Clear-sky obs. error on land (K)
1	50.3	0.33		
2	52.8	0.21		
3	53.246	0.21		
4	53.596	0.22	0.48	0.55
5	54.4	0.20	0.48	0.5
6	54.94	0.20	0.6	0.61
7	55.5	0.21	0.72	0.73
8	57.290344	0.22	1.2	1.24
9	89	0.10		
10	165.5	0.18		
11	176.311	0.18	3.88	4.47
12	178.811	0.18	4.22	4.62
13	180.311	0.25	3.52	4.01
14	181.511	0.25	3.67	3.96
15	182.311	0.35	3.71	3.91
16	325.15 ± 1.2	0.47		
17	325.15 ± 2.4	0.39		
18	325.15 ± 4.1	0.32		
19	325.15 ± 6.6	0.25		

TABLE 3 Description of the frequencies onboard the instrument of the EPS-Sterna satellites, the NeΔT used to simulate the observations (corresponding to the raw NeΔT divided by a factor three) and the clear-sky observation errors used over oceans and land to assimilate the observations. Channels 1 to 8 are temperature sounding channels, channels 9 and 10 are window channels, channels 11 to 15 are humidity sounding channels and channels 16 to 19 are humidity and ice-cloud sounding channels.

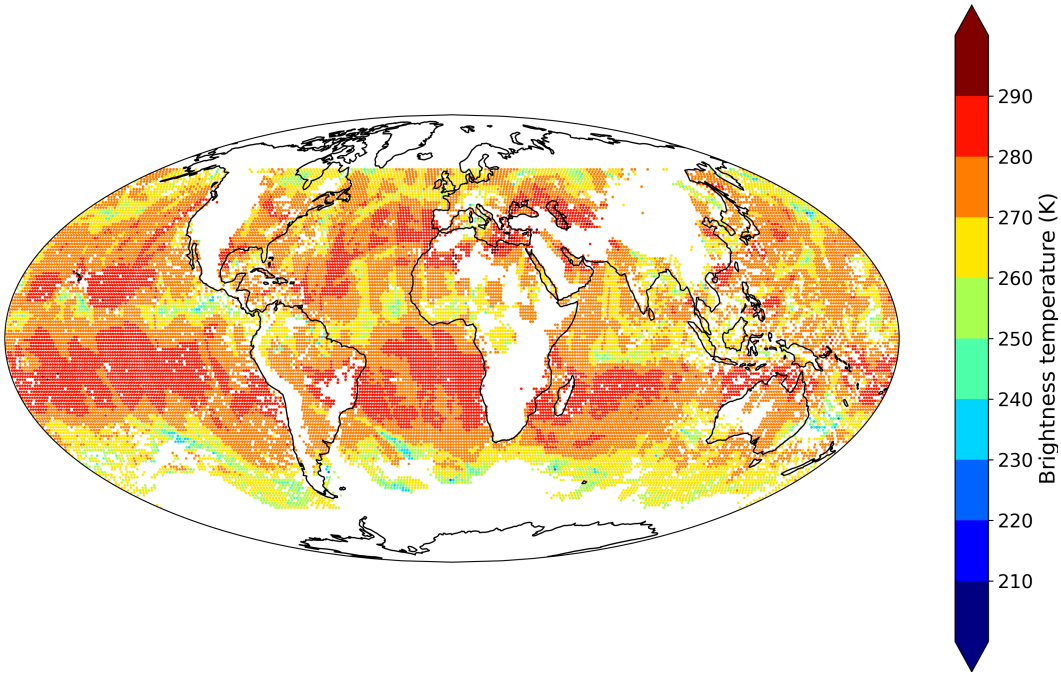


FIGURE 3 Simulated observation values of EPS-Sterna OP3-6SAT scenario for band 11 (176.311 GHz) between 17/08/2021 21:00 UTC and 18/08/2021 03:00 UTC. Flagged data are not shown.

Name	Description
CONTROL OSSE	Simulated observations assimilated as presented in table 1
NO METOP-B OSSE	CONTROL OSSE - METOP-B
OP3-3SAT	CONTROL OSSE + OP3-3SAT
OP2-4SAT	CONTROL OSSE + OP2-4SAT
OP3-4SAT	CONTROL OSSE + OP3-4SAT
OP3-6SAT	CONTROL OSSE + OP3-6SAT
OP4-8SAT	CONTROL OSSE + OP4-8SAT
CONTROL REAL	Real observations assimilated as presented in table 1
NO METOP-B REAL	CONTROL REAL - METOP-B

TABLE 4 Description of the various data assimilation experiment runs. A detailed presentation of the observations used is presented in section 2.4.

time and we estimate that the time of dissemination of Sterna data would be of one hour and a half on average.

3 | EXPERIMENTAL SET-UP AND VERIFICATION METHODS

3.1 | Experimental set-up

The framework presented in section 2 is used to run various experiments, which are presented in table 4. First of all, a control experiment is run with the reference observing system and without any EPS-Sterna observations. Then, a Metop-B denial experiment (denial of the scatterometer, AMSU-A, MHS, IASI, GNSS-RO bending angles onboard Metop-B) is run, and used to validate the OSSE framework. Finally, five experiments are run with the reference observing system, on top of which the different EPS-Sterna scenarios are added (see section 2.5).

All these experiments are conducted during the two periods, but the Metop-B denial experiment, which is conducted only during the December to February period. The exact duration of the periods are: between the 14 August 2021 and the 30 October 2021, and between the 5 December 2021 and the 28 February 2022.

In addition, the control experiment as well as the Metop-B denial experiments are also run with the assimilation framework using real observations presented in section 2.3. This will allow the inter-comparison of the forecast error growth of the control experiments and of the impacts of Metop-B on forecast errors (see section 4.1).

3.2 | Verification methods

In order to (i) validate the OSSE framework by comparing the impact of a Metop-B denial experiment computed using this framework and real observations, and (ii) measure the impact of the various EPS-Sterna scenarios compared to the control experiment, various metrics were set up and are described in sections 3.2.1, 3.2.2 and 3.2.3 below. For all these metrics, the forecasts produced by the experiments presented in table 4 are compared to the nature run, which provides the real state of the simulated atmosphere. The different metrics are computed over different geographical areas, such as the globe (all grid points are used), the Northern Hemisphere (between latitudes 20° and 90°), the Tropics (between latitudes -20° and 20°) or the Southern Hemisphere (between latitudes -90° and -20°). These metrics are

computed using either one or both of the periods combined. When the two periods are used, results are computed using 164 forecasts up to +102 h forecast lead time.

3.2.1 | Standard deviation of the forecast error

The standard deviation of forecast error is a traditional metric in NWP and enables a comparison of the forecast skill of different experiments regarding temperature, wind speed, humidity and geopotential height. For each +102 h forecast at 00:00 UTC, the differences between the forecasts of each experiment and the nature run are computed. There is one forecast error value per grid point for all vertical levels (between 1000 hPa and 100 hPa) and all forecast lead times. From the distribution of error values, a forecast error standard deviation is computed using all the grid points over a given geographical area. This produces a forecast error standard deviation as a function of the vertical level and the forecast lead time for various geophysical variables: temperature, wind speed, relative humidity and geopotential height, as shown in figure 4. This forecast error standard deviation is then compared between a given EPS-Sterna experiment and the control experiment: the results are then represented as relative forecast error standard deviations, in percent, as shown in figures 6 or 8. These results are computed using data interpolated over a regular grid of 0.25 degrees of resolution. Statistical significance is computed using a bootstrap test at the 99% confidence level (Efron, 1992).

3.2.2 | FSOI-like metric

This FSOI-like metric enables the production of an integrated score, which sums up forecast error in an energy norm expressed in J kg^{-1} m^{-2}. This norm is based on the same norm as the one used in Forecast Sensitivity to Observations Indices (FSOi), which is widely used among meteorological centers (Eyre, 2021), with adaptations to use the nature run as the reference instead of an analysis. It is computed using the energy norm proposed by Ehrendorfer et al. (1999) and has already been used with the ARPEGE model, as presented in Chambon et al. (2023). This norm is computed using the difference between the forecast of the experiments and the truth of the nature run, and by summing this forecast error for four different variables: surface pressure, temperature, wind speed and relative humidity, with normalisation factors specific to each variable. It is evaluated using all the grid points, all vertical levels between 1000 hPa and 1 hPa and for various forecast lead times. The results are then compared between each EPS-Sterna experiment and the control OSSE experiment, and thus the relative energy norm results, in percent, are presented. These results are computed using raw data, over the native ARPEGE tilted stretched grid, thus a weighting of the data depending on the location of the grid point has been set up. Error bars correspond to an uncertainty of the mean at the 99% confidence level. An example of the output of such a metric can be found in figure 11.a.

3.2.3 | Precipitation scores

Precipitation is not taken into account in the two previously presented metrics. Yet, the assimilation of microwave sounders can have a significant impact on precipitation forecast skill of a model, as presented in Geer et al. (2017). Thus, Fractions Skill Scores (FSS) have been computed, using the methodology developed by Roberts and Lean (2008). With this metric, one computes the difference between the daily precipitation forecasts and the nature run reality but taking into account a neighbourhood to avoid the double penalty issue associated with misplaced rainy systems. A neighbourhood of size of 5×5 pixels (0.5 degrees per 0.5 degrees) is used, associated with an accumulation threshold of 3 mm. Relative difference between the FSS of the OP3-6SAT experiment and the control OSSE experiment are

computed: the results presented are relative FSS, in percent. These results are computed using data interpolated over a regular grid of 0.1 degree resolution. Statistical significance is computed using the bootstrap test at the 99% confidence level. An example of the output of such a metric can be found in figure 14.

4 | RESULTS

4.1 | Validation of the OSSE framework

Using the two frameworks presented above (the OSSE framework with simulated observations and the OSE framework with real observations), we conducted two comparisons: a comparison of forecast error growths in the two control runs as well as a comparison of forecast degradations from the denial of Metop-B.

Figures 4 and 5 show the comparison between the two control experiments using the forecast error standard deviation metric presented in section 3.2.1. One can see that the trends and structures of error growth are similar for wind (Figures 4.b vs 4.f and Figures 5.b vs 5.f) and geopotential height (Figures 4.d vs 4.h and Figures 5.d vs 5.h), over the two hemispheres. Some differences can be noticed for temperature (Figures 4.a vs 4.e and Figures 5.a vs 5.e) and relative humidity (Figures 4.c vs 4.g and Figures 5.c vs 5.g) for which the error growth is lower, especially near the surface. In general the forecast error standard deviations for the CONTROL REAL experiment are higher. This can be explained by at least two factors. First, the CONTROL REAL experiment is validated with the ECMWF analysis which has its own errors. This can inflate the standard deviation compared to the one of the CONTROL OSSE which is compared to the "truth". Secondly, the simplification of the OSSE framework related to the surface data assimilation might explain the smaller error growth which can be seen close to the surface, in particular for temperature forecasts (Figures 4.a vs 4.e and 5.a vs 5.e). The underestimation of forecast error growth in the lower layers has implications on the final results of calibration of observations shown in section 2.4, leading for instance to an optimistic match of MHS first-guess departure standard deviation channels 4 and 5 but could be improved in a future study as will be discussed in section 5.

Figures 6 and 7 show the relative degradations of forecast error standard deviation from the denial of one Metop satellite, for the Northern and Southern Hemispheres and for three different key variables: wind speed at 250 hPa, geopotential height at 500 hPa and temperature at 850 hPa. redNote that the calibration result already validates the impact of observing systems in the short range (+6 h), therefore, this complementary validation focuses on longer ranges starting from +24 h.

Figures 6 and 7 show that for the two geographical zones, and the three variables, the denial of one Metop degrades the forecasts by about 2% up to a 4-day range. For most of the forecast lead times and three variables, one can also see a good agreement in the magnitude of impact of the denial of one Metop between the OSSE and the experiment with real observations. The degradation of forecast errors is, in some cases, slightly larger in the OSSE than the degradation induced in the real observations framework. The OSSE framework seems to slightly overestimate the impact of the observations, in particular in the short range, which could partly be the result of some observations being incorrectly calibrated. However, the overall trends on each of these three key variables are highly similar. It strengthens the confidence in the OSSE framework, and its capability to reproduce the impacts of a real observing system, with maybe a small overestimation of the impacts.

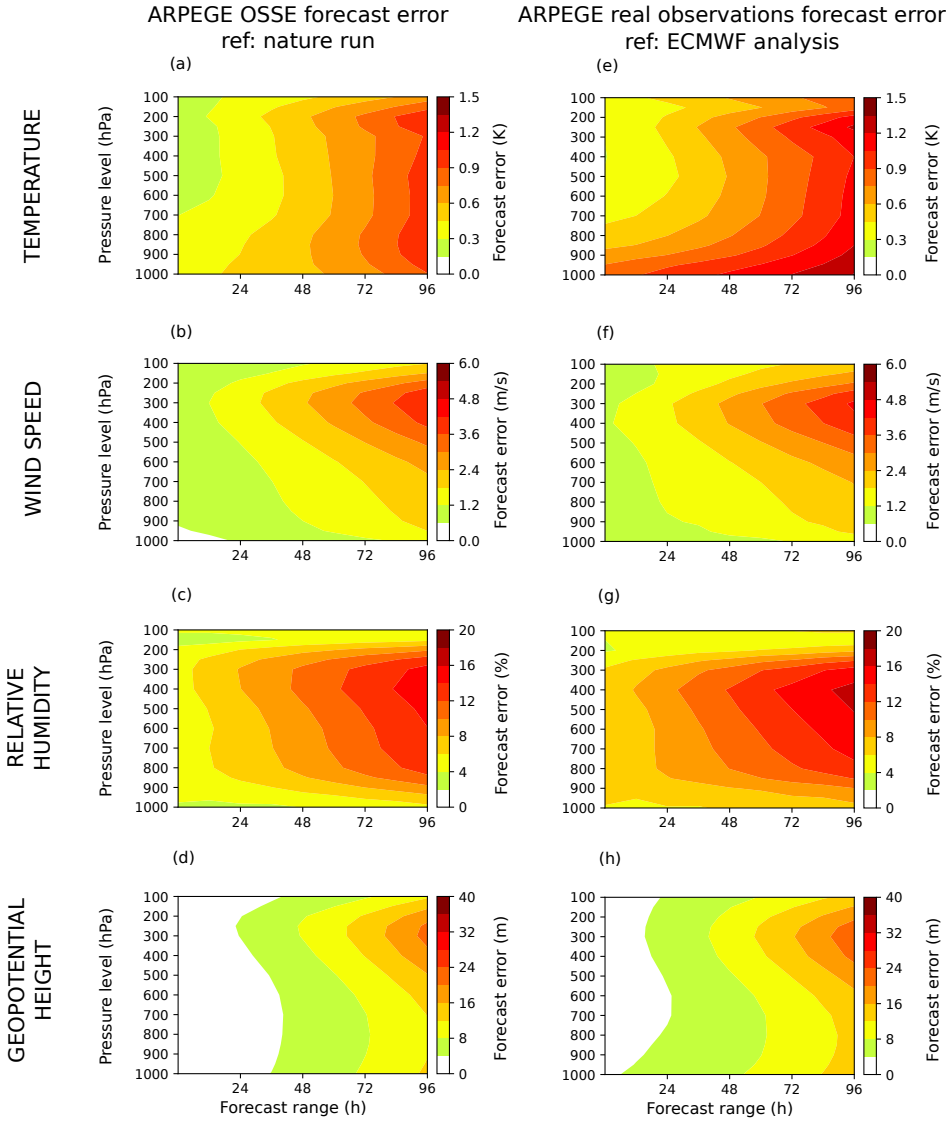


FIGURE 4 Absolute forecast error standard deviation of the CONTROL OSSE experiment (left column) with the Nature Run as reference and of the CONTROL REAL experiment (right column) with the ECMWF analysis as reference. The first row shows statistics for temperature, the second row for winds, the third row for relative humidity and the fourth row for geopotential height. Each score is presented as a function of forecast range in hours on the x-axis and as a function of vertical level in hPa on the y-axis. The statistics have been computed for the Northern Hemisphere, on a period ranging from 5 December 2021 to 28 February 2022.

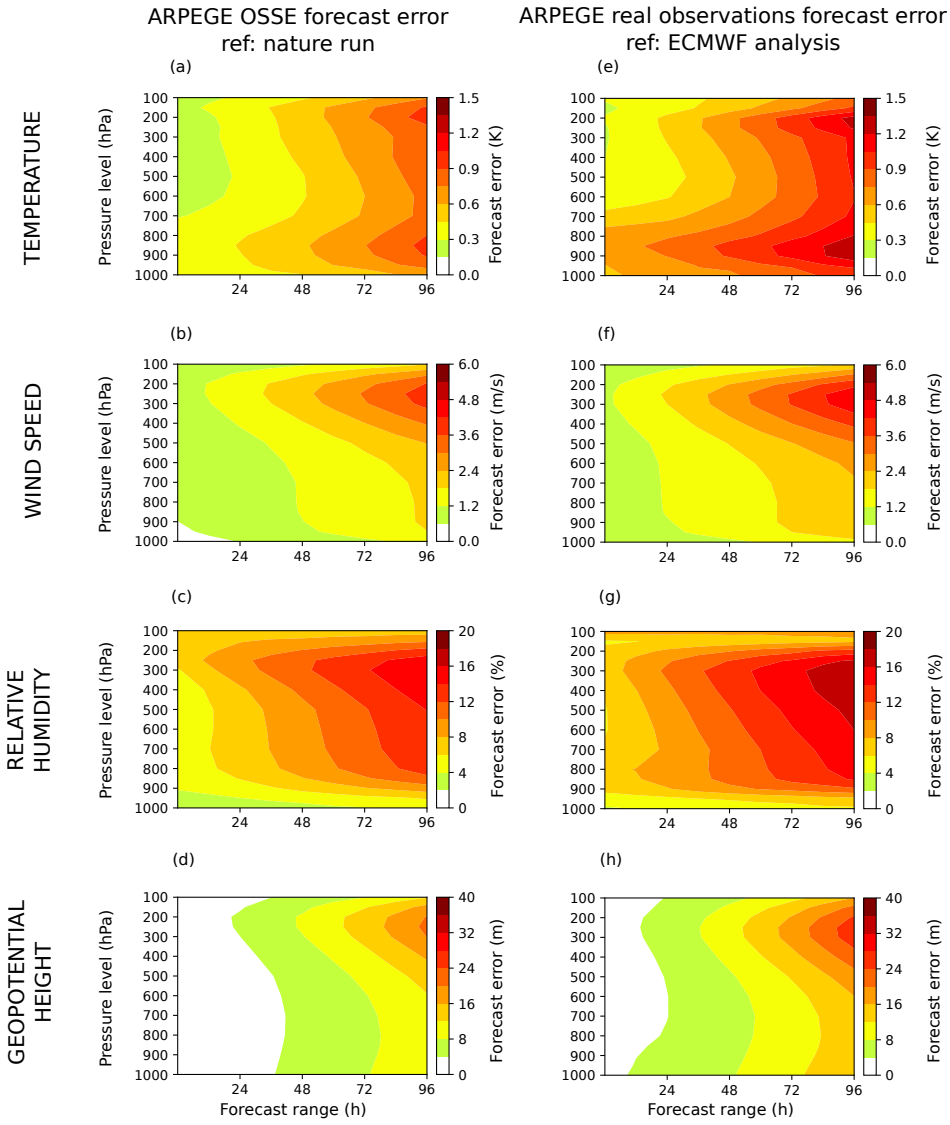


FIGURE 5 Same as figure 4, but over the Southern Hemisphere.

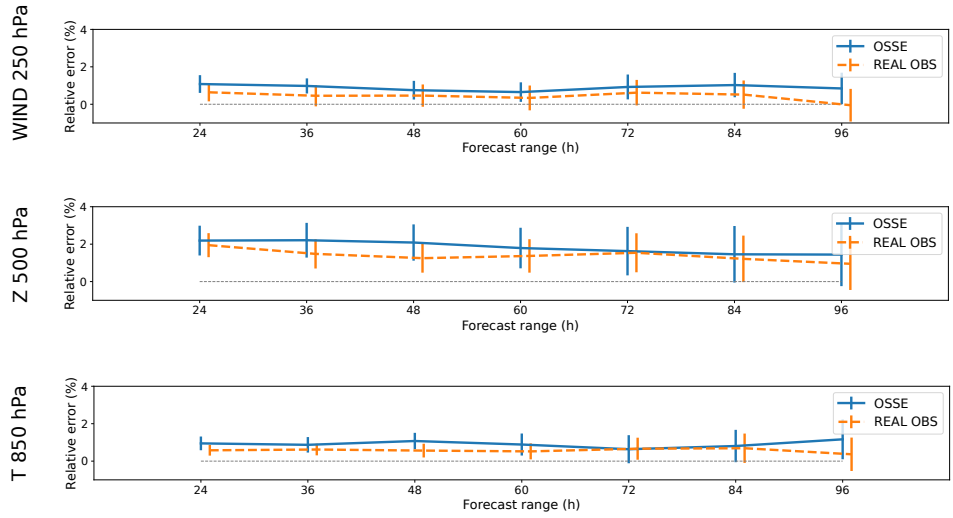


FIGURE 6 Intercomparison of impacts of a Metop-B denial within the OSSE framework (nature run as reference) and within a data assimilation experiment with real observations (ECMWF analyses as reference), over the Northern Hemisphere and the period ranging from 5 December 2021 to 28 February 2022. For each framework, the relative standard deviation of the Metop-B denial experiment is compared to the CONTROL experiment, and results are displayed in percent. The scores are displayed as a function of forecast range in hours, for one given vertical level: 250 hPa for wind speed (top), 500 hPa for geopotential height (middle) and 850 hPa for temperature (bottom). The error bars correspond to uncertainty in the mean at the 99% confidence level.

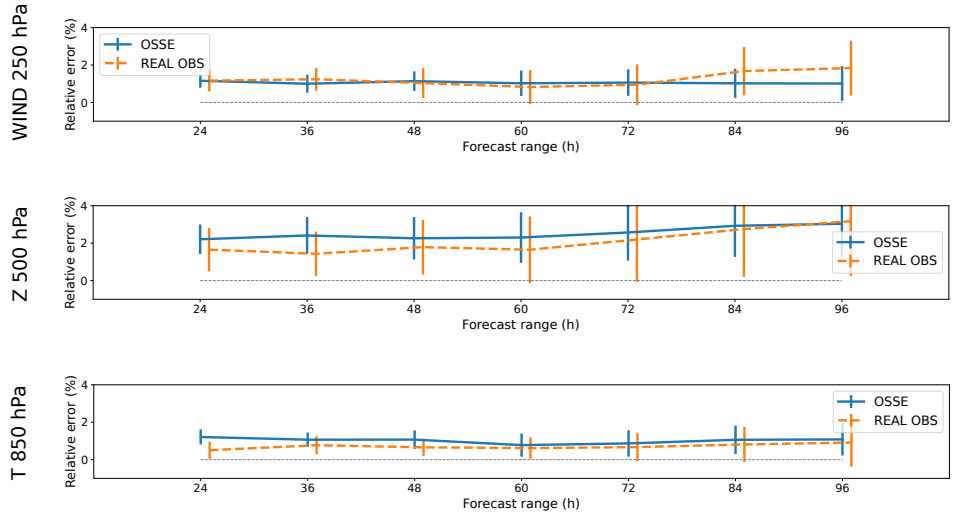


FIGURE 7 Same as figure 6, showing an intercomparison of impacts of a Metop-B denial within the OSSE framework and within a data assimilation experiment with real observations, but over the Southern Hemisphere.

4.2 | Impact of EPS-Sterna on standard deviation of the forecast error

The scenarios OP2-4SAT, OP3-6SAT and OP4-8SAT presented in section 3.1 are compared to the nature run using the standard deviation of the forecast error presented in section 3.2.1.

Figure 8 presents the results for the OP3-6SAT scenario, over three different geographical zones: the Northern Hemisphere, the Tropics and the Southern Hemisphere. Figure 10 presents the results for the globe for three different scenarios: OP2-4SAT, OP3-6SAT and OP4-8SAT.

Figures 8.i, 8.j, 8.k and 8.l show a positive and significant reduction of forecast error for the the OP3-6SAT scenario in the Southern Hemisphere, for the four variables (temperature, wind speed, relative humidity and geopotential height), at all vertical levels up to +96 h. The error reduction goes up to 6% for geopotential height at short range around 300 hPa. Figures 8.a, 8.b, 8.c and 8.d present a smaller but still positive and significant impact of the OP3-6SAT scenario in the Northern Hemisphere, for the four variables, at all vertical levels up to 48h. The impact goes up to 3.5% on relative humidity at short range around 400 hPa. In the Tropics (figures 8.e, 8.f, 8.g and 8.h), the impact is positive, significant and can go up to 96h at some levels. A degradation can also be noticed on humidity (figure 8.g) in the short range around 500 hPa. This negative signal might come from several sources but is likely related to tropical convection and the way the EPS-Sterna observations are assimilated within these meteorological scenes. Several tests have been performed (e.g. denial of different channels, alignment of convection schemes between the assimilation and the nature run). The main conclusion which was drawn is that this degradation comes from the usage of the humidity channels and further investigations will be needed to prevent this effect from occurring when using this kind of observations. Note that this negative signal has already been observed in the literature with real observations when assessing the impact of the SAPHIR humidity sounder onboard the Megha-Tropiques satellite in the ECMWF model, as presented in Chambon and Geer (2017).

Figures 9 and 10 allow the comparison of the impact of the number of satellites and orbital planes of the EPS-Sterna scenarios at the global scale. In figure 9, results for the globe are overlaid for the short range (+12 h) and on figure 10, impacts up to +96 h are presented for the globe.

At +12 h forecast lead time in figure 9, one can see that the forecast errors are reduced for the four variables, by 1% to 4%. For temperature, the incremental improvements derived by adding additional satellites can be clearly seen from 150 hPa down to 700 hPa. A similar conclusion can be drawn for winds and geopotential heights with an enhanced impact when adding satellites (the importance of well distributed satellites will be further discussed later in section 4.3). In the case of humidity, the improvements are less homogeneous in the vertical at the global scale and are the result of the combination of the impact on different geographical areas, as previously shown, with improvements in the Northern and Southern Hemisphere and degradations between 400 hPa and 700 hPa in the Tropics.

In the longer range, figure 10 shows that the differences between scenarios persist in time at the global scale. As for the OP3-6SAT scenario, the two others have positive and significant impacts for all vertical levels up to +96 h, for all variables presented. For all variables, the impact increases when a new pair of satellites is added (Temperature : figures 10.a, 10.e, 10.i ; Winds: figures 10.b, 10.f, 10.j ; Relative humidity : figures 10.c, 10.g, 10.k ; Geopotential height: figures 10.d, 10.h, 10.l). For instance, the temperature forecast error reduction goes up to 0.8% at +48 h forecast lead time for the OP2-4SAT scenario (figure 10.a), 1.0% for the OP3-6SAT scenario (figure 10.e) and 1.3% for the OP4-8SAT scenario (figure 10.i).

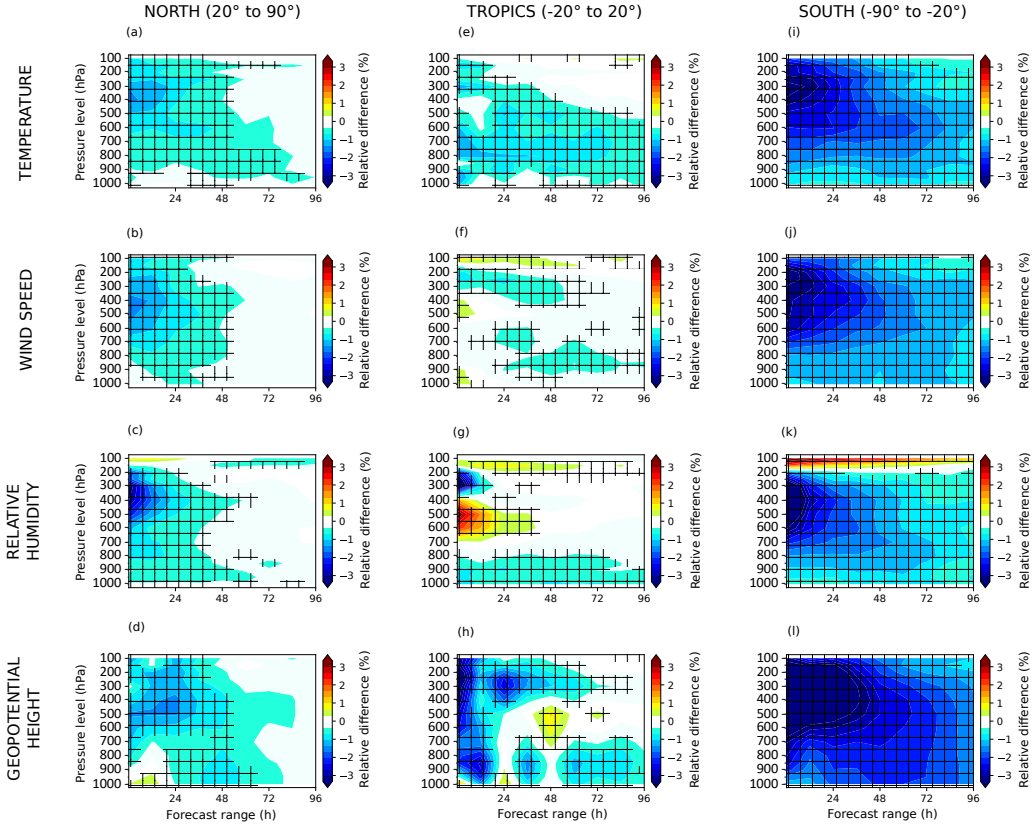


FIGURE 8 Relative difference of standard deviation of the forecast error for the OP3-6SAT scenario compared to the CONTROL scenario, in percent. Results are presented for three different geographical zones (Northern Hemisphere, Tropics and Southern Hemisphere) and four different variables (temperature, wind speed, relative humidity and geopotential height), using 164 forecasts. Each score is presented as a function of forecast range in hours on the x-axis and as a function of vertical level in hPa on the y-axis. The reference used to compute the forecast error is the nature run. Hatched areas indicate statistical significance at the 99% confidence level computed with the bootstrap method.

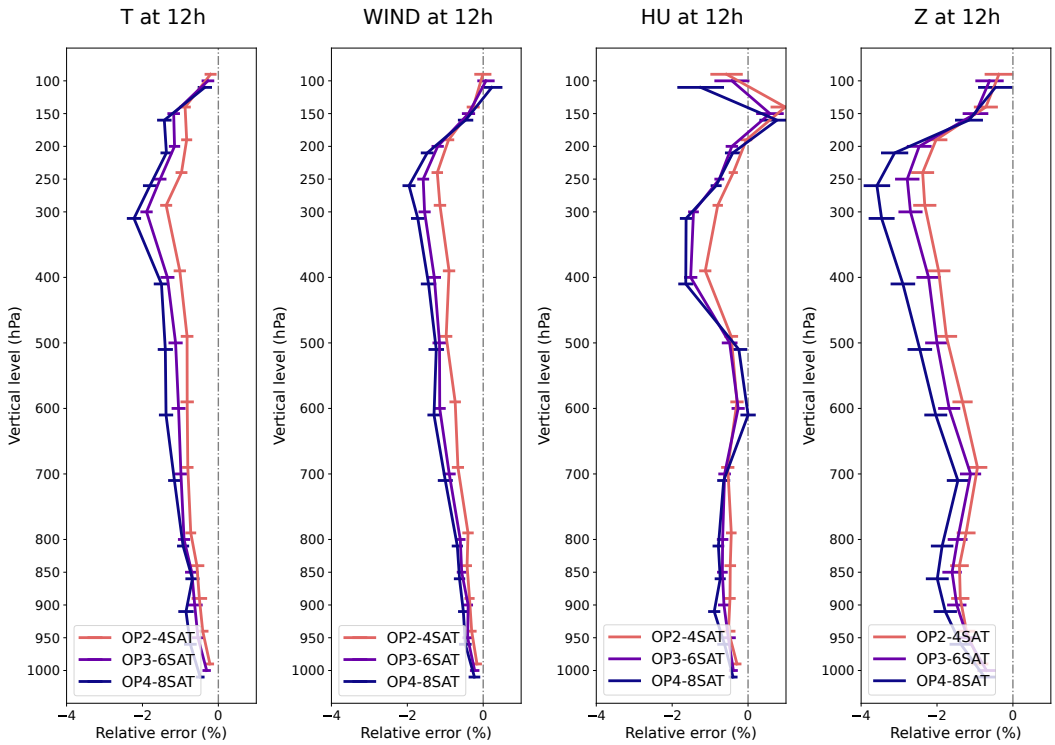


FIGURE 9 Relative difference of standard deviation of the forecast error for the OP2-4SAT, OP3-6SAT and OP4-8SAT scenarios compared to the CONTROL scenario, in percent, for the +12 h forecast lead time. Results are presented for the globe for four different variables (temperature, wind speed, relative humidity and geopotential height), using 164 forecasts. Each score is presented with the relative difference in percent on the x-axis and the vertical levels in hPa on the y-axis. The reference used to compute the forecast error is the nature run. The error bars correspond to uncertainty in the mean at the 99% confidence level.

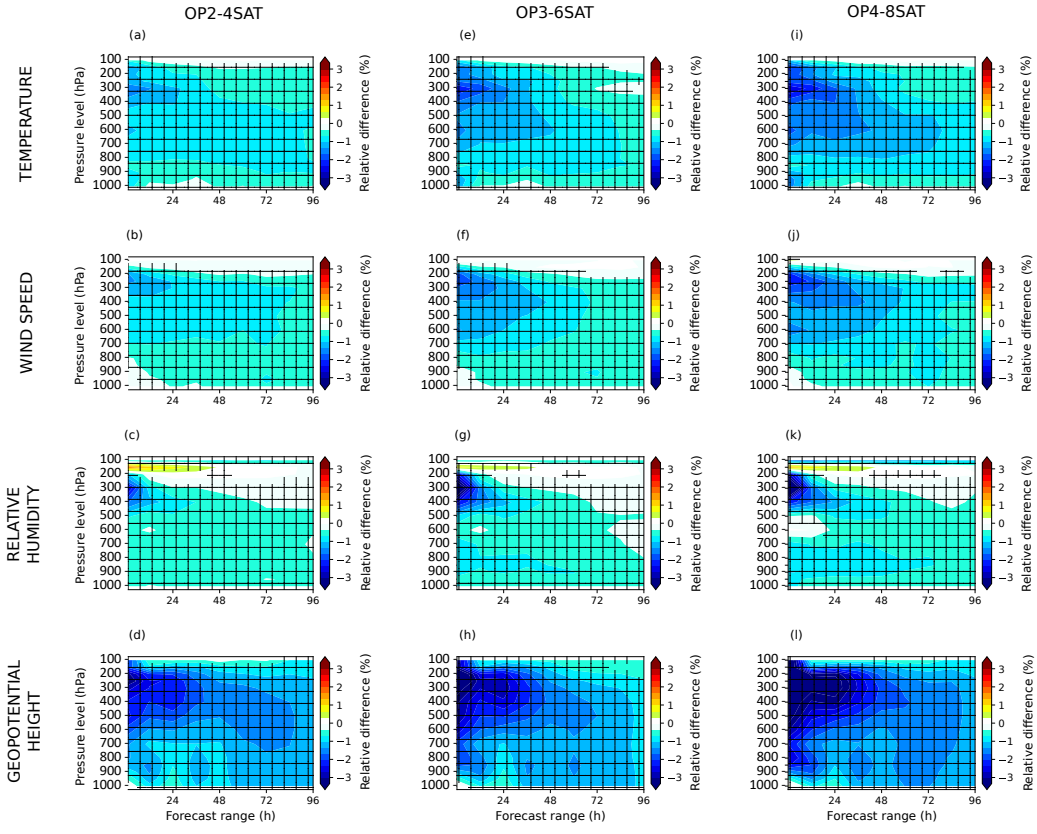


FIGURE 10 Relative difference of standard deviation of the forecast error for the OP2-4SAT, OP3-6SAT and OP4-8SAT scenarios compared to the CONTROL scenario, in percent. Results are presented for the globe for four different variables (temperature, wind speed, relative humidity and geopotential height), using 164 forecasts. Each score is presented as a function of forecast range in hours on the x-axis and as a function of vertical level in hPa on the y-axis. The reference used to compute the forecast error is the nature run. Hatched areas indicate statistical significance of the differences of scores at the 99% confidence level computed with the bootstrap method.

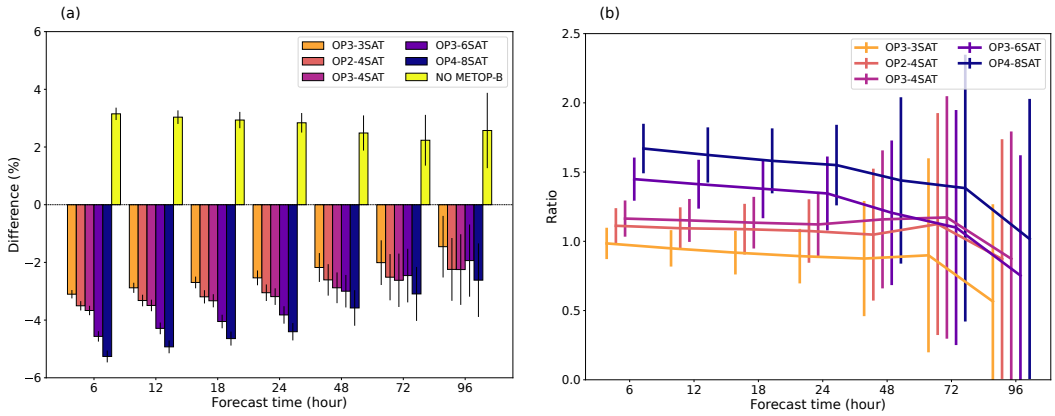


FIGURE 11 Relative difference of energy norm of the forecast error for the OP3-3SAT, OP2-4SAT, OP3-4SAT, OP3-6SAT, OP4-8SAT and NO METOP-B OSSE scenarios compared to the CONTROL scenario, in percent. The left figure (a) represents energy norm for each scenario and the right figure (b) represents the absolute value of the ratio between the relative difference of energy norm derived from adding EPS-Sterna and denying Metop-B. Results are presented as a function of forecast range in hour, over the Southern Hemisphere, using 164 forecasts. The reference used to compute the forecast error is the nature run. The error bars correspond to uncertainty on the mean at the 99% confidence level.

4.3 | Impact of EPS-Sterna on FSOI-like metric

The scenarios OP3-3SAT, OP2-4SAT, OP3-4SAT, OP3-6SAT and OP4-8SAT presented in section 3.1 are compared to the nature run using the FSOI-like metric presented in section 3.2.2.

Figure 11.a shows the impact of the scenarios presented in 3.1 over the Southern Hemisphere. It can be noticed that the impacts of the EPS-Sterna scenarios are positive and significant up to +96 h: the decrease of energy norm representing the fact that the forecast error has decreased as shown in the previous section. No saturation can be observed, as the more satellites, the larger the impact, at least in the short range. The impact of a Metop-B denial is of the same order of magnitude compared to the impact of some EPS-Sterna scenarios. To get a precise understanding of the differences and similarities between these two impacts, the relative differences of energy norm of the EPS-Sterna scenarios presented in figure 11.a are divided by the relative difference of energy norm of the NO METOP-B OSSE experiment presented in the same figure. In figure 11.b, one can note that the impact of the OP2-4SAT scenario in the short range is about 1.2 times the impact of a Metop-B satellite, 1.5 for the OP3-6SAT scenario and 1.7 for the OP4-8SAT scenario. Note that this comparison should be taken with caution as it focuses on one specific application of Metop (NWP) among other ones (e.g. sea surface wind speed, atmospheric composition, air quality monitoring, etc.).

Figure 12.a shows the impact of the scenarios presented in 3.1 over the Northern Hemisphere. It is noticeable that the impact of the EPS-Sterna OP3-6SAT scenario is lower than over the Southern Hemisphere, yet still positive and significant up to +48 h. Such a difference of impact could be explained by the smaller number of conventional observations in the Southern Hemisphere, which tend to less constrain the atmospheric analysis compared to that of the Northern Hemisphere. This difference of impact between the Southern Hemisphere and the Northern Hemisphere has already been observed with operationally assimilated microwave sounding observations in Duncan et al. (2021), with a ratio of impact between the two hemispheres similar to the one found between figures 11 and 12. Figure 12.b

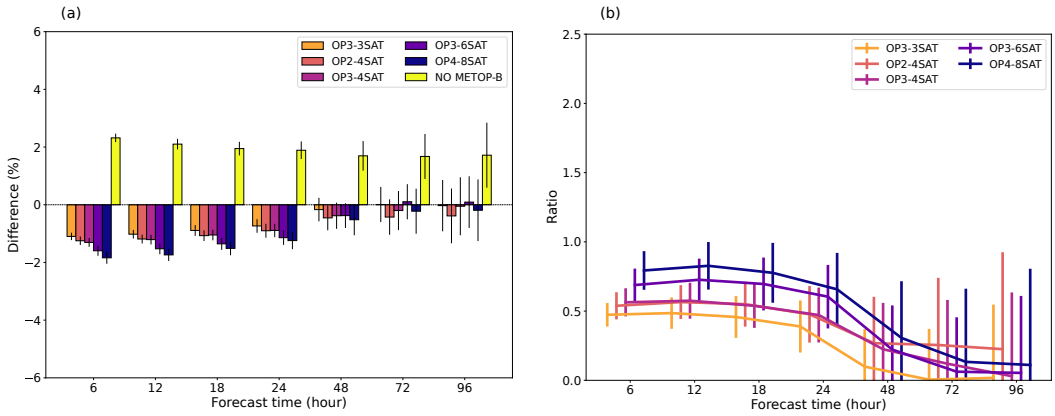


FIGURE 12 Same as figure 11, showing relative difference of energy norm of the forecast error, but in the Northern Hemisphere.

presents the ratios of relative difference of energy norm between the EPS-Sterna scenarios and the NO METOP-B OSSE experiment. One can evaluate the ratio of impact: in the Northern Hemisphere, the impact of the OP2-4SAT scenario in the short range is about 0.5 times that of a Metop-B satellite, the OP3-6SAT is about 0.7 times and the OP4-8SAT is about 0.8 times. The differences of impact ratios (EPS-Sterna / Metop) between the Northern Hemisphere and the Southern Hemisphere indicates that the magnitude of improvement is decreasing faster in the Northern Hemisphere when adding new observations, likely due to the higher density of other observation kinds.

In order to measure the impact of the distribution of the satellites in the orbital planes, two scenarios have been compared. The energy norm of the forecast error for the OP2-4SAT scenario and the OP3-4SAT have been computed, then the relative difference is taken between these two scenarios, with the OP2-4SAT being the reference. This produces figure 13, which shows whether the addition of an orbital plane, with the total number of satellites being held constant, has a significant impact on forecast error. One can observe that the OP3-4SAT scenario has a significant and positive impact at short range compared to the OP2-4SAT scenario. Thus, a better distribution of observations acquired from a given number of satellites over more orbital planes has a significant impact on forecast error reduction at short range.

4.4 | Impact of EPS-Sterna on precipitation

Figure 14 presents a comparison of the Fractions Skill Score for the OP3-6SAT experiment and the CONTROL experiment, as presented in section 3.2.3. The impact of the OP3-6SAT scenario is positive and significant up to +96 h over the Southern Hemisphere, and less important but still positive and significant up to +48 h over the Northern Hemisphere. In the Tropics, one can observe a significant degradation of the forecast skill of the model at short range (which matches with the degradation of the humidity field in the short range mentioned before), and a significant improvement at longer range. When averaging statistics over the globe, the impact of the OP3-6SAT scenario is positive and significant up to +96 h. These scores are overall quite consistent compared to the scores shown before on humidity forecasts. The same scores have been computed for the other EPS-Sterna scenarios (not shown). Overall, the impacts of these scenarios on precipitation forecasts over the Northern Hemisphere, the Tropics and the Southern Hemisphere are very similar. The differences among the scenarios are not statistically significant. This reflects the fact

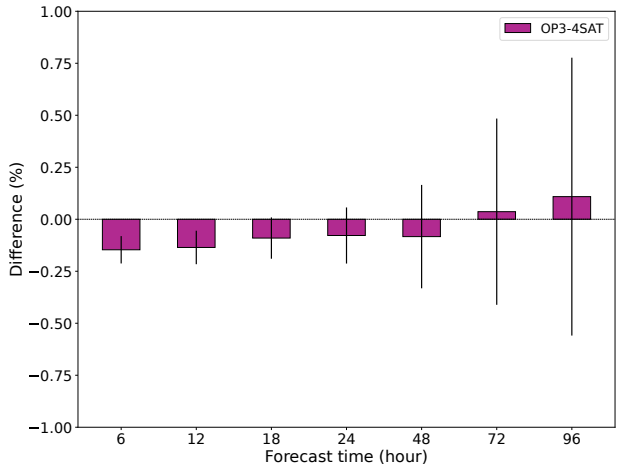


FIGURE 13 Relative energy norm of the forecast error for the OP3-4SAT scenario compared to the OP2-4SAT scenario, in percent. Results are presented as a function of forecast lead time in hours, for the globe, using 164 forecasts. The reference used to compute the forecast error is the nature run. The error bars correspond to uncertainty of the mean at the 99% confidence level.

that while microwave data definitely have an impact on precipitation forecasts (Geer et al., 2017), this geophysical variable remains harder to improve than others.

5 | CONCLUSIONS

The objective of this study is to provide an overview of the potential benefits of a constellation of microwave sounders, named EPS-Sterna, which would be similar to the AWS satellite planned to be launched in 2024. For that purpose, an OSSE framework has been set up, following the steps presented as follows:

- A nature run has been computed over two four-months periods.
- Synthetic observations have been produced, and the perturbations added to these simulated observations have been tuned to match first-guess departure statistics of real observations.
- A data assimilation framework for these simulated observations has been set up and validated by comparing the forecasts produced to real forecasts, through a Metop denial experiment. In particular, the selected setup avoids the "identical twin problem".
- Metrics have been set up to measure the impact of the EPS-Sterna scenarios on various variables, at different forecast ranges, vertical levels and for different geographical zones.

The main findings of this study are presented below:

- A validation of the OSSE framework constructed in this study was performed. The Metop-B denial experiments conducted with simulated observations and with real observations produce similar results on key meteorological

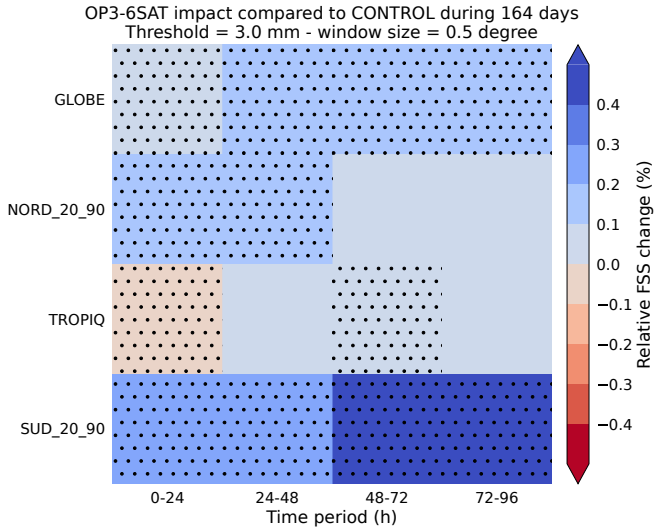


FIGURE 14 Relative Fractions Skill Score (FSS) of 24h accumulated precipitation for the OP3-6SAT scenario compared to the CONTROL experiment, in percent. Results are presented as a function of forecast range in hour, for various geographical zones, using 164 forecasts. The reference used to compute the forecast error is the nature run. Dotted areas indicate statistical significance at the 99% confidence level computed with the bootstrap method.

variables, with a small overestimation of the impact of the observations at some forecast ranges.

- Every EPS-Sterna scenario tested has a positive and significant impact on temperature, wind speed, relative humidity, geopotential height and precipitation forecasts.
- For all metrics used, the impact of the EPS-Sterna scenarios is larger in the Southern Hemisphere than in the Northern Hemisphere, yet always positive. In the Tropics, the impact can be slightly negative in the short range for humidity, but positive at a longer range.
- An enhanced impact of EPS-Sterna has been demonstrated when more satellites are added to the constellation.
- No saturation of the forecast error reduction when increasing the number of microwave sounders has been observed.
- The OP3-6SAT scenario is comparable to 1.7 Metop-B in the Southern Hemisphere and 0.7 in the Northern Hemisphere.
- In the short range, a positive impact was detected when adding an orbital plane with a constant number of satellites, thus demonstrating the value of a better distribution of the satellites among orbital planes.

It is important to point out that the comparison between the impacts of a Metop satellite and the EPS-Sterna constellation made in this study only covers the improvements of weather forecasts in the troposphere. It certainly does not cover the full range of applications that can benefit from Metop observations and should by no means be understood as proposing that one observing system can replace the other.

Overall, one of the main conclusions that can be drawn from this study is that launching the EPS-Sterna constellation can greatly and significantly improve forecasts up to +96 h in the troposphere. In a future study, it will also be interesting to document the impacts of EPS-Sterna in the stratosphere.

As a perspective, the framework presented could be improved on several aspects in future studies:

- The fraternal twin aspect of the OSSE framework could be avoided by using an alternative nature run from another global model. One possibility would be for instance to use the 9 km ECMWF Cubic Octahedral (O1280) grid Nature Run (Hoffman et al., 2018). This may indeed increase the realism of forecast error growth. However, this nature run is available every 3 hours compared to the one tailored for this study, available every 30 minutes. Using this alternative nature run would therefore affect the realism of simulated observations by using the same state of the atmosphere for observations within 3-hour time slots. By not taking advantage of consecutive observations with time differences shorter than 3 hours, the realism of the impact assessment of the EPS-Sterna constellation may be changed. Therefore the cumulative effect of improving the realism of forecast error growth but degrading the realism of the EPS-Sterna constellation simulation would be interesting to assess in a future study.
- The surface data assimilation system could be calibrated and run in addition to the atmospheric one. This would increase the realism of the OSSE system, thus improving the realism of the EPS-Sterna simulations. This would be important for assessing the impact of future observations characterised by an information content closer to the surface than in the EPS-Sterna concept.
- The simulation of several observation types could be improved, in particular the simulation of Atmospheric Motion Vectors which have been simulated in the OSSE from the nature run winds but their consistency with the nature run meteorological conditions could be improved (Errico and Prive, 2018; Errico et al., 2020).
- Climatological background errors have been computed and remain unchanged across the different scenarios which are sub-optimal compared to the usage of daily computed background errors as in operations.
- The reference observing system was defined to mimic the observing system of 2030 by removing existing satellites for which replacements are not foreseen. However, the exact list of missions which will provide valuable data for NWP is hard to predict, in particular due to the emergence of the private sector in the domain of Earth observation. Some planned missions were also not taken into account in the baseline, like the infrared hyperspectral sounders onboard geostationary satellites (e.g. the future InfraRed Sounder onboard Meteosat Third Generation-S). Therefore, it would be interesting to revisit these results of EPS-Sterna impact with a more up-to-date baseline observing system.
- More validation of the OSSE framework in the Tropics would be needed to improve the understanding of the short range forecast degradations found with the EPS-Sterna constellation.
- Regarding the assimilation of the EPS-Sterna satellites, optimisations could be performed, such as having specific cloud predictors for temperature channels (Duncan et al., 2022) instead of using the same one for both humidity and temperature observations, or such as actively using the 325 GHz channels either as cloud predictors or for data assimilation.

Acknowledgements

EUMETSAT is acknowledged for funding the work presented in this paper under the contract number EUM/COS/LET/21/12484. The authors thank Emma Turner for providing the RTTOV-SCATT coefficients used in this study. Clément Laval is acknowledged for the initial work he conducted, which enabled this study to be successful. The MWS/AWS Science Advisory Group members are acknowledged for the useful comments they made on the study which helped improving the framework. The anonymous reviewers are acknowledged for the insightful and constructive reviews they made, which significantly improved the quality of this manuscript. Michael Glinton is acknowledged for proof-reading the manuscript.

Data Availability Statement

The data produced are available upon request to the authors.

references

Batté, L. and Déqué, M. (2016) Randomly correcting model errors in the ARPEGE-Climate v6. 1 component of CNRM-CM: applications for seasonal forecasts. *Geoscientific Model Development*, **9**, 2055–2076.

Berre, L. and Desroziers, G. (2010) Filtering of background error variances and correlations by local spatial averaging: A review. *Monthly Weather Review*, **138**, 3693–3720.

Berre, L., Varella, H. and Desroziers, G. (2015) Modelling of flow-dependent ensemble-based background-error correlations using a wavelet formulation in 4D-Var at Météo-France. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2803–2812.

Bormann, N., Healy, S., Lean, K. and Lonitz, K. (2023) Predicting the forecast impact of potential future observing systems. *ECMWF Newsletter*, **174**, 12–17.

Bougeault, P. (1985) A simple parameterization of the large-scale effects of cumulus convection. *Monthly Weather Review*, **113**, 2108–2121.

Bouyssel, F., Berre, L., Bénichou, H., Chambon, P., Girardot, N., Guidard, V., Loo, C., Mahfouf, J.-F., Moll, P., Payan, C. et al. (2022) The 2020 Global Operational NWP Data Assimilation System at Météo-France. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)*, 645–664.

CGMS (2022) CGMS Baseline: Sustained contributions to the observing of the Earth system, space environment and the Sun. *Tech. rep.*, Coordination Group For Meteorological Satellites.

Chambon, P. and Geer, A. J. (2017) All-sky assimilation of Megha-Tropiques/SAPHIR radiances in the ECMWF numerical weather prediction system. *Tech. rep.*, European Centre for Medium-Range Weather Forecasts.

Chambon, P., Mahfouf, J.-F., Audouin, O., Birman, C., Fourrié, N., Loo, C., Martet, M., Moll, P., Payan, C., Pourret, V. et al. (2023) Global Observing System Experiments within the Météo-France 4D-Var Data Assimilation System. *Monthly Weather Review*, **151**, 127–143.

Coopmann, O., Fourrié, N., Chambon, P., Vidot, J., Brousseau, P., Martet, M. and Birman, C. (2023) Preparing the assimilation of the future MTG-IRS sounder into the mesoscale NWP AROME model. *Quarterly Journal of the Royal Meteorological Society*.

Duncan, D. I., Bormann, N., Geer, A. J. and Weston, P. (2022) Assimilation of AMSU-A in all-sky conditions. *Monthly Weather Review*, **150**, 1023–1041.

Duncan, D. I., Bormann, N. and Hólm, E. V. (2021) On the addition of microwave sounders and numerical weather prediction skill. *Quarterly Journal of the Royal Meteorological Society*, **147**, 3703–3718.

Duruisseau, F., Chambon, P., Guedj, S., Guidard, V., Fourrié, N., Taillefer, F., Brousseau, P., Mahfouf, J.-F. and Roca, R. (2017) Investigating the potential benefit to a mesoscale NWP model of a microwave sounder on board a geostationary satellite. *Quarterly Journal of the Royal Meteorological Society*, **143**, 2104–2115.

Efron, B. (1992) Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, 569–593. Springer.

Ehrendorfer, M., Errico, R. M. and Raeder, K. D. (1999) Singular-vector perturbation growth in a primitive equation model with moist physics. *Journal of the Atmospheric Sciences*, **56**, 1627–1648.

Errico, R. M., Carvalho, D., Privé, N. C. and Sienkiewicz, M. (2020) Simulation of atmospheric motion vectors for an observing system simulation experiment. *Journal of Atmospheric and Oceanic Technology*, **37**, 489–505.

Errico, R. M. and Prive, N. C. (2018) Some general and fundamental requirements for designing observing system simulation experiments (OSSEs). *Tech. rep.*, World Meteorological Organization.

Errico, R. M., Yang, R., Privé, N. C., Tai, K.-S., Todling, R., Sienkiewicz, M. E. and Guo, J. (2013) Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. *Quarterly Journal of the Royal Meteorological Society*, **139**, 1162–1178.

ESA (2023a) Description of the AWS satellite on the ESA web page. `http://web.archive.org/web/20230612083610/https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/Arctic_Weather_Satellite`. Accessed on June 6th, 2023.

— (2023b) Description of the AWS sensor on the ESA web page. `http://web.archive.org/web/20230612083734/https://www.esa.int/Applications/Observing_the_Earth/Meteorological_missions/Arctic_Weather_Satellite/The_instrument`. Accessed on June 6th, 2023.

Eyre, J. (2021) Observation impact metrics in NWP: A theoretical study. Part I: Optimal systems. *Quarterly Journal of the Royal Meteorological Society*, **147**, 3180–3200.

Geer, A., Baordo, F., Bormann, N., Chambon, P., English, S., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. and Lupu, C. (2017) The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, **143**, 3189–3206.

Geer, A. J. and Bauer, P. (2011) Observation errors in all-sky data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **137**, 2024–2037.

Geer, A. J., Bauer, P., Lonitz, K., Barlakas, V., Eriksson, P., Mendrok, J., Doherty, A., Hocking, J. and Chambon, P. (2021) Bulk hydrometeor optical properties for microwave and sub-millimetre radiative transfer in RTTOV-SCATT v13. 0. *Geoscientific Model Development*, **14**, 7497–7526.

Guedj, S., Guidard, V., Ménétrier, B., Mahfouf, J.-F. and Rabier, F. (2014) Future benefits of high-density radiance data from MTG-IRS in the AROME fine-scale forecast model. *Tech. rep.*, Eumetsat Fellowship final report, Météo-France & CNRS/CNRM-GAME.

Hoffman, R. N., Malardel, S., Maciel, P., Peevey, T. R., Partain, P., Finley, S., Atlas, R. M., Isaksen, L., Wedi, N., Cucurull, L. et al. (2018) The New ECMWF Cubic Octahedral (O1280) Nature Run. In *AGU Fall Meeting Abstracts*, vol. 2018, A13C–01.

Masutani, M., Schlatter, T. W., Errico, R. M., Stoffelen, A., Andersson, E., Lahoz, W., Woollen, J. S., Emmitt, G. D., Riishøjgaard, L.-P. and Lord, S. J. (2010) Observing system simulation experiments. *Data Assimilation: Making sense of observations*, 647–679.

Privé, N., Errico, R. and Akkraoui, A. E. (2022) Investigation of the potential saturation of information from Global Navigation Satellite System radio occultation observations with an observing system simulation experiment. *Monthly Weather Review*, **150**, 1293–1316.

Roberts, N. M. and Lean, H. W. (2008) Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136**, 78–97.

Roehrig, R., Beau, I., Saint-Martin, D., Alias, A., Decharme, B., Guérémy, J.-F., Voldoire, A., Abdel-Lathif, A. Y., Bazile, E., Belamari, S. et al. (2020) The CNRM global atmosphere model ARPEGE-Climat 6.3: Description and evaluation. *Journal of Advances in Modeling Earth Systems*, **12**, e2020MS002075.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A. et al. (2018) An update on the RTTOV fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, **11**, 2717–2737.

593 Stark, J. D., Donlon, C. J., Martin, M. J. and McCulloch, M. E. (2007) OSTIA: An operational, high resolution, real time, global
594 sea surface temperature analysis system. In *Oceans 2007-europe*, 1–4. IEEE.

595 Stoffelen, A., Marseille, G., Bouttier, F., Vasiljevic, D., De Haan, S. and Cardinali, C. (2006) ADM-Aeolus Doppler wind lidar
596 observing system simulation experiment. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric*
597 *sciences, applied meteorology and physical oceanography*, **132**, 1927–1947.

598 Tiedtke, M. (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly weather*
599 *review*, **117**, 1779–1800.