SOFF
Systematic Observations Financing Facility

ECMWF
EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

ECMWF SOFF Impact Experiments

November 2026

Table of contents

Executive Summary 3
1. Context 6
2. Scope and methodology of Impact Experiments 7
2.1 Ensemble of Data Assimilations (EDA) and Ensemble Forecasting (ENS) methodology 7
2.2 Assimilation of simulated data 9
3. Data coverage 10
4. Results from the phase 3 experiments 12
4.1 SOFF medium-range impact on global and hemispheric scales 12
4.2 Impacts by region (localized impacts) 13
4.3 Impact of SOFF expansion on the Atlantic Main Development Region (MDR) 17
5. Concluding remarks 19
6. Next steps 19
7. References 19

Executive Summary

Accurate weather forecasts save lives and livelihoods, and improved networks of weather observations increase forecast accuracy. Yet vast areas of the world are under-observed by surface-based, in situ measurements. To assess the benefits of closing these gaps and guide SOFF investments, the SOFF Steering Committee commissioned a series of science-based impact studies and experiments, carried out by the European Centre for Medium-Range Weather Forecasts (ECMWF) in coordination with WMO and the SOFF Secretariat.

ECMWF is an intergovernmental organization that is both a research institute and a 24/7 operational service producing global weather predictions for its 35 Members and Co-operating States as well as for a large user base consisting of governmental and inter-governmental agencies worldwide as well as research and commercial organizations.

ECMWF's research was undertaken in three phases.

During **phase 1,** ECMWF reviewed the scientific literature on the role of observations in improving forecast accuracy (Decision 9.2).

Phase 1 findings:

- Existing studies show that **surface-based observations significantly impact forecast accuracy.**
- **Surface-based observations create local, regional and global benefits.** When a new observation is used its impact will, at first, be local, but as weather systems move the impact moves with them so that each day the area benefiting from the original observation becomes larger.
- **Stations in data sparse regions have a much larger individual contribution to the forecast accuracy** than any station in a well observed region, underlying the importance of global coverage by the Global Observing System.
- **Surface-based observations remain essential in a satellite-rich observing system**. In the Northern Hemisphere, surface-based in-situ observations used at operational forecasting centers have a greater impact on forecasts than satellite observations.
- **The importance of surface-based observations grows as forecasting technology evolves.** Surface-based in situ information has an even bigger impact in new, machine-learning based systems than in traditional approaches.
- **Investing in closing the surface-based observation gaps is highly beneficial** and a key element necessary to improve forecast accuracy.

In **phase 2**, ECMWF designed and ran eight tailored observation network "scenarios" simulating the impact of expanding Global Basic Observing Network (GBON) surface, upper-air and marine stations (Decision 11.2). These assessed the forecast improvements from adding observations in Least Developed Countries (LDCs), Small Island Developing States (SIDS), Lower Middle-Income Countries (LMICs), and closing regional in-situ observation gaps, such as in Africa and the Pacific. ECMWF quantified how new

observations reduce uncertainty in forecasts covering the next 12 hours – providing rigorous and comparable evidence of the potential of these investments.

Phase 2 findings:

- **More observations and data deliver greater impact:** Forecast accuracy improves in proportion to the number of new observations added.
- **Greatest benefits for Africa:** The largest improvements are seen from adding observations in Africa, reducing forecast uncertainty by over 30% in high-impact regions.
- **Investments in the Pacific SIDS matter:** Forecast uncertainty reduced by up to 20% in this data-sparse and important region for global weather prediction.
- **Data from upper air stations deliver the largest benefits in the tropics:** Upper-air observations are especially important in the tropics for constraining atmospheric profiles.
- **SOFF expansion to all Official Development Assistance (ODA) countries gives the biggest return**. The expansion scenario to Lower-Middle Income Countries and all ODA countries shows the largest improvement in forecast accuracy.
- **Local investments generate global benefits:** Weather does not respect borders, so improved observations strengthen forecast improvements across regions well beyond borders.

Phase 2 showed that it would be desirable to extend the impact assessment into the medium range. Therefore, in **phase 3**, ECMWF developed a methodology to assess the impact of additional observations on medium-range forecasts, defined by WMO as from day 3 to day 10, and applied it to two observation scenarios from phase 2. The first scenario assessed the impact of closing the GBON data gap in Least Developed Countries (LDCs) and Small Island Developing States (SIDS). The second scenario covered a broader group of countries, namely all Official Development Assistance (ODA)-eligible countries and SIDS. ECMWF also extended the experiments from one month to five months, covering June 1 to October 31, 2024, to achieve statistically robust results in the medium range.

Phase 3 results :

- **More observations create bigger impact:** Closing the GBON data gap in all ODA-eligible and SIDS countries has a larger impact in the medium range than closing the data gap in LDCs and SIDS only for most of the variables and regions.
- **Medium-range forecast accuracy improves in both the northern and southern hemisphere extra-tropics[1]:** Closing the GBON data gap in all ODA-eligible and SIDS countries improves northern-hemisphere extra-tropical forecast accuracy, for temperature, wind and humidity and typically to around day 5.

[1] The region of the hemispheres outside the tropical zone, including middle and high latitudes.

- Covering LDCs and SIDS improves forecast accuracy by around one day less, extending to around day 4 for many variables.
- **Largest improvements in the tropics:** Most significant forecast accuracy improvements on upper-air temperature, humidity and wind are found up to 7 day forecasts for both scenarios studied. Surface temperature and low-level winds forecasts are also improved out to day 6 when averaged for this area.
- **Significant gains for Africa:** Given the regional distribution of new observations in both scenarios, clear, statistically significant forecast improvements are seen for Africa extending into the medium range, often beyond day 5.
- **Pacific forecast improvements:** Improved forecast accuracy from temperature, wind and humidity for the Tropical Western Pacific region are statistically significant out to around day 5.
- **African observing network improvements are important for Atlantic hurricanes preparedness:** Improvements originating in Africa spread westward into the North Atlantic for longer forecasts. More specifically, they improve medium-range forecasts for a key meteorological area known as the Main Development Region and the impact is demonstrable to day 5. Around 60% of Atlantic tropical cyclones develop from African Easterly Waves in this region. The results indicate that forecasts of upper-level winds are improved by a 4–5-hour lead time at day 5. In other words, African observation network improvements contribute to improved Atlantic tropical cyclone preparedness, with downstream relevance for vulnerable regions such as the Caribbean.

These ECMWF science-based impact studies and experiments provide the strongest scientific evidence to date that targeted investments in GBON infrastructure in under-observed regions dramatically boost forecast accuracy, both locally and globally, providing a clear case for scaled-up SOFF investments.

ECMWF Impact Experiments

1. Context

The SOFF Steering Committee, through Decision 6.8, endorsed and approved funding for SOFF Impact Reports. Impact refers to the improvement in forecast skill resulting from increased investment in observations. It has been proven that the value of improved forecast skill translates directly to safety and numerous socio-economic benefits[2].

WMO engaged the European Centre for Medium-Range Weather Forecasts (ECMWF), one of the World Meteorological Centres and member of the SOFF Advisory Board, to prepare the assessments for the reports. The study was initially conducted in two phases. The first phase (Decision 9.2) provided a literature review and assessment of current knowledge on forecast skill improvement and impact of improved surface and upper-air observation networks. Phase one also identified the scenarios for SOFF-tailored further research and experiments in phase two.

In phase two, impact studies were performed for eight tailored SOFF investment scenarios. These numerical experiments provided evidence that additional GBON observations improve short-range forecasts, particularly in the regions where the density of new observations was high. For example, large potential forecast impacts in Africa were presented. However, it was also recognised that further studies would be required to estimate the impact of these observation scenarios on medium-range forecasts.

In phase three, ECMWF developed a new assessment approach and extended the duration of experimentation for two SOFF scenarios to assess their potential impact on medium-range forecast scores, where the medium range is assumed to cover day 3 to day 10 in this study.

This document summarises the results from the new experiments completed during phase three.

All monitoring and prediction of weather start from the collection and global exchange of observations. This data provides the only direct source of information about the atmosphere. Weather is inherently global, and to understand and predict weather and climate patterns, observations covering the entire globe need to be made available to the global monitoring and prediction model systems. The WMO Integrated Global Observing System (WIGOS) comprises both satellite- and surface-based observations. Spaceborne observations offer global coverage, capturing data from every part of the Earth. However,

[2] Mark R. Rosenzweig and Christopher R. Udry: Assessing the Benefits of Long-Run Weather Forecasting for the Rural Poor: Farmer Investments and Worker Migration in a Dynamic Equilibrium Model; NATIONAL BURAU OF ECONMIC RESEARCH Working Paper No. 25894, Cambridge, MA, May 2019.

the information that satellites provide is "indirect" and it usually requires the application of sophisticated inverse methods to retrieve useful information about the atmospheric state from the raw measurements. Calibrating and supplementing satellite data with directly measured surface and upper-air data is essential in this context.

Despite the global coverage provided by the variety of satellite observing systems used in operational numerical weather prediction (NWP), phase one of this study confirmed that surface-based data, including both surface and upper-air in situ observations, remain a crucial component of WIGOS. In fact, the combined surface based in-situ dataset used at operational forecasting centers still has a larger forecast impact in the northern hemisphere than any individual satellite observing system. Further, in addition to improving the initial conditions required to produce accurate forecasts, they also have important additional roles in both forecast verification and in constraining the bias corrections applied to satellite radiance measurements.

Surface-based observations directly measure critical weather parameters such as temperature and humidity, but are limited by the availability of observation stations, particularly in remote and underdeveloped areas. Currently, there are large data gaps in surface-based weather observation networks, negatively affecting the quality of weather forecasts globally. Closing these data gaps is essential for the world to be better prepared, to better understand and therefore more effectively adapt to a changing climate.

2. Scope and methodology of Impact Experiments

2.1 Ensemble of Data Assimilations (EDA) and Ensemble Forecasting (ENS) methodology

To assess the potential impact of additional surface-based observations in the medium range (day 3 to day 10), ECMWF developed a new approach, combining both Ensemble Data Assimilation (EDA) and Ensemble Forecasting System (ENS) experiments. The EDA component of this approach has been applied by researchers to estimate the impact of a number of new satellite observing systems (e.g., Lean *et al.*, 2025), and it was used in phase 2. This is an ensemble method that provides a *theoretical* estimate of the "error bars" for the NWP analyses (initial conditions) and short-range forecasts, both with and without the new simulated datasets. In short, the EDA used in this study is composed of ten *perturbed* NWP forecasting systems, plus an unperturbed control member. These are run in parallel, with random noise added to each perturbed member. If the random noise is chosen *appropriately* (as discussed in Healy et al., 2024), the EDA "spread" can be used to estimate the error bars of the analyses and short-range forecasts. The "spread" is simply the standard deviation of an ensemble of atmospheric states about the mean of the ensemble. If the EDA spread is reduced as a result of assimilating a new dataset, this indicates that the new dataset is expected to improve the statistical uncertainty – or error bar – of the analysis and/or short-range forecasts.

The EDA system is designed to provide uncertainty estimates for the analysis or short-range forecasts, and it produces the perturbations that represent initial state uncertainty used in the ECMWF's ensemble forecasting system (ENS), which subsequently provides uncertainty information for medium-range forecasts.

In phase 3, ECMWF has extended EDA experiments out to day 5 and performed new ENS experiments out to day 15. This combination of experiments provides a more complete picture for medium-range impact and enables the consistency of the EDA and ENS results to be assessed.

This study is novel because it represents the first use of the ENS system by ECMWF to predict the impact of a new dataset on medium-range forecasts. However, the limitations of the current methodology should be recognised. In particular, ensemble methods where random noise is deliberately introduced into the NWP systems, like the EDA and ENS, can be sensitive to sampling errors. These are a result of a finite ensemble size (in this case 10 perturbed members) and a limited experimental period. To partially mitigate this, ECMWF has extended the experimental period from 1 month to 5 months. Nevertheless, in some cases these sampling errors can still produce counter-intuitive noisy results, particularly when assessing relatively small areas on spatial maps. This noise is usually smoothed out in large scale spatial averages, but this limitation should be noted when interpreting some results provided in this study.

The investigation presents statistical features derived over a defined and relatively short five-month period. Consequently, individual high-impact synoptic events in Europe, such as Saharan dust outbreaks or strong precipitation events associated with energy input from data-scarce regions in Africa, may not be adequately represented during this period. It can be assumed that, for such events, additional information from data-scarce regions could provide valuable input to forecasting systems over Europe.

The phase 2 scenarios extended in the phase 3 study (Table 1) were:

- Baseline (Control): Current observing system (shown in Figure 1 based on June 2023).
- SC 1: Closing GBON data gap in Least Developed Countries (LDC) *and* Small Island Developing States (SIDS)
- SC 3: Closing GBON data gap in all Official Development Assistance (ODA) Countries *and* Small Island Developing States (SIDS)

These scenarios represent significant surface and upper-air network changes, but SC 3 is more comprehensive (See Table 2).

The baseline (control) scenario assimilates all satellite and in-situ observations that were used operationally in June 2023 and includes no simulated data. The combined number of real in situ and satellite observation values assimilated in the baseline control experiment is greater than 30 million per day. The new simulated observations (surface [S], upper air [UA]) were added to the baseline for SC 1 and SC 3, as described in Table 1.

Table 1. The eight observation scenarios originally studied in phase 2. In phase 3, SC 1 and SC 3 experiments (highlighted rows) were extended to cover the period June 1 to October 31, 2023.

Scenario	Country groupings	Data type
SC 1	LDCs + SIDS	S + UA
SC 2	LDCs + SIDS + LMICs	S + UA
SC 3	All ODA + SIDS	S + UA
SC 4	LDCs + SIDS	S + UA + Marine
SC 5	FCS	S + UA
SC 6	Pacific	S + UA
SC 7	Africa	S + UA
SC 8	All ODA + SIDS	UA only

[ODA = Official Development Assistance; LDC = Less Developed Country; LMIC = Low- and Middle-Income Countries; SIDS = Small Island Developing State; FCS = Fragile and Conflict Affected Situations] [S=Surface Stations; UA=Upper Air stations]

Table 2. The scenario definitions, observation numbers and expected annual compliance cost. The highlighted rows are extended in phase 3.

Scenarios		Number of additional surface stations (improved+new)	Number of additional upper-air stations (improved+new)	Number of marine stations
SC 1	LDCs+SIDS | S+UA	617	135	
SC 2	LMICs+LDCs+SIDS | S+UA	1069	194	
SC 3	All ODA+SIDS | S+UA	1709	259	
SC 4	LDCs+SIDS | S+UA+marine	617	135	145
SC 5	FCS | S+UA	576	110	
SC 6	Pacific | S+UA	117	19	
SC 7	Africa | S+UA	703	137	
SC 8	All ODA+SIDS | UA only		259	

2.2 Assimilation of simulated data

Simulated observations were generated from ECMWF's high resolution, operational four-dimensional variational[3] (4D-Var) analyses, using the same simulation code as in phase 2.

[3] 4D-Var is the system used at ECMWF to combine a short-range forecast with observations to produce the best possible estimate of the current state of the atmosphere. This then provides the initial conditions for a new forecast.

The temporal frequency of the simulated data met GBON requirements for surface (hourly) and radiosondes (six-hourly, to simulate the impact of both ascending and descending radiosonde measurements which are now used operationally). Noise added to the simulated data is assumed to be both spatially and temporarily uncorrelated. It has been demonstrated that the statistical characteristics of simulated data generated in this way are similar to real observations.

Simulated datasets for this study are subjected to the same quality control procedures as real observations when they are assimilated. As in phase 2, the only exception is for surface winds over land, which are not currently assimilated in the operational ECMWF system but have been tested in the study.

3. Data coverage

Figure 1 and Figure 2 illustrate the in-situ observation locations of both the real (orange) and simulated (teal) surface and upper-air datasets used in this study, for SC 1 and SC 3, respectively. The bounding boxes will be referred to later in section 4.2. These figures again emphasise the complementary, gap-filling nature of the proposed observation network changes and they help explain why this new data is likely to have significant impact on NWP forecasts. The value of a new observation depends on the prior information at that location. Gap-filling observations in data-sparse regions—where the prior from a previous forecast is less certain—therefore tend to deliver the largest benefits for NWP. The plots also illustrate that SC 3 is a more comprehensive change and this is why it produces larger reductions in forecast uncertainty, as demonstrated in phase 2.

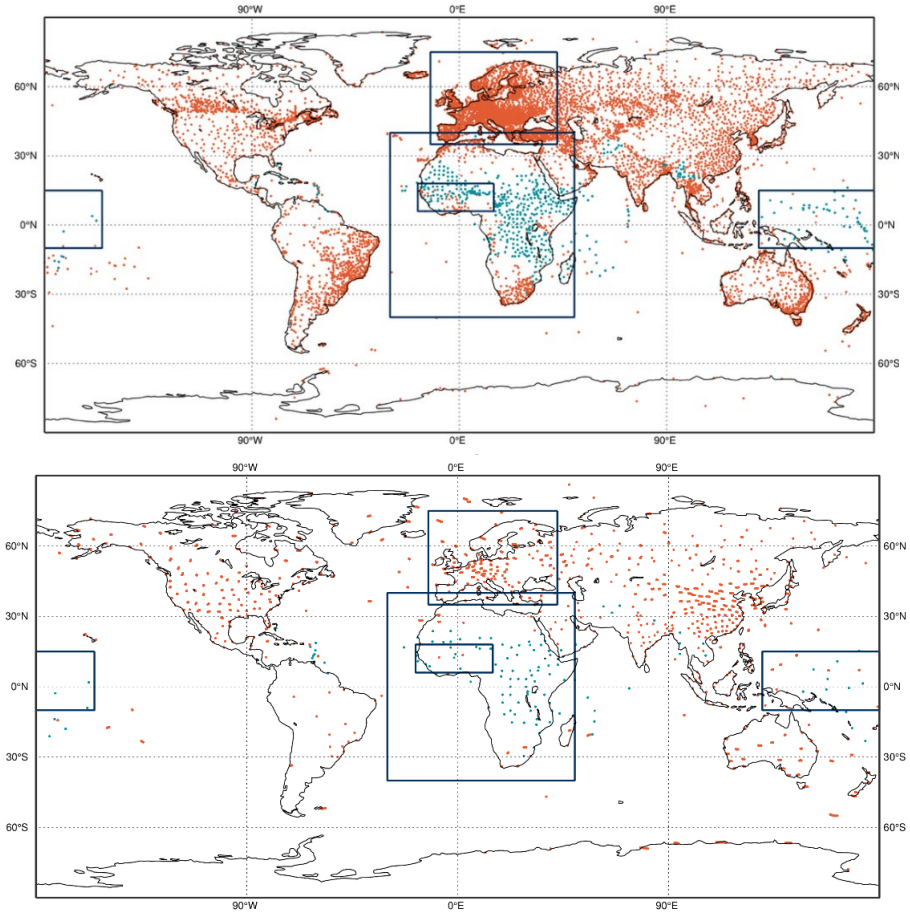


Figure 1. Spatial coverage of SYNOP surface pressure (upper) and SONDE (lower) measurements used in a 12-hour assimilation window on June 1, 2023, for the ***SC 1*** experiment. The real observations are shown as orange and the simulated observations added to the baseline experiment are shown in teal. The boxes will be referred to later in the report.

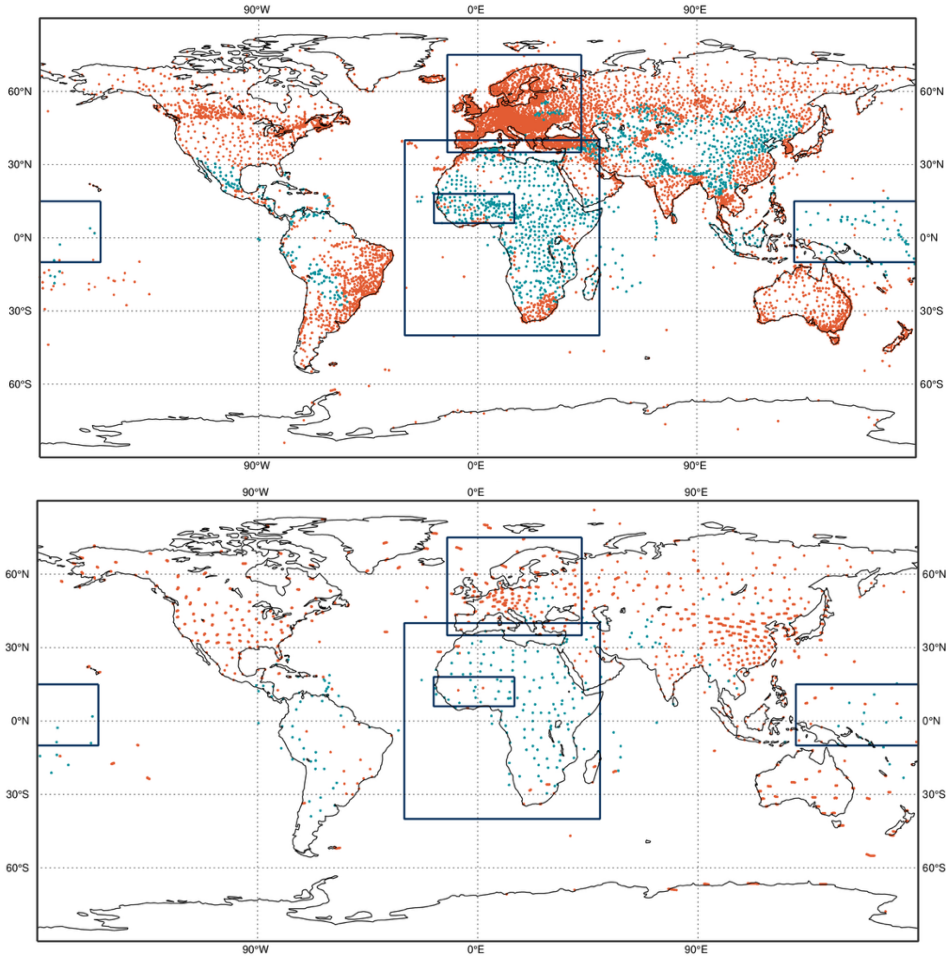


Figure 2. Spatial coverage of SYNOP surface pressure (upper) and SONDE (lower) measurements used in a 12-hour assimilation window on June 1, 2023, for the ***SC 3*** experiment. The real observations are shown in orange and the simulated observations added to the baseline experiment are shown in teal. The boxes will be referred to later in the report.

4. Results from the phase 3 experiments

4.1 SOFF medium-range impact on global and hemispheric scales

Clear, statically significant forecast impact in the tropics to day 7

Figure 3 is an example of the medium-range spread changes obtained with SC 1 and SC 3. It shows the vertical profile of spread reduction for zonal wind ("u") at day 3 (upper) and day 5 (lower), when averaged globally and when split into the northern hemisphere extra-tropics (latitude > 20°N), tropics (20°N < latitude > 20°S) and southern hemisphere (20°S > latitude). The solid lines represent the EDA results, and the corresponding ENS results on the 850, 700 and 250 hPa pressure levels are denoted by crosses. The error bars represent the two-sided 95% confidence interval (half-width) of the mean difference, estimated using the Student's t distribution, computed from the time-series of spread differences. In this work, spread reductions are interpreted by ECMWF as a statistically significant positive impact relative to the control experiment, if the corresponding error

bars do not cross the zero line. Similarly, increases in spread do not represent a forecast degradation of the forecast quality if the error bar associated with the spread change crosses the zero line.

When assessing medium-range impacts, it should be recognized that it has become increasingly difficult to demonstrate the impact of good observing system changes on medium-range forecast error statistics. This is because the existing global observing system is gradually becoming more robust and resilient. As a result, long periods of experimentation are often now required to achieve statistical significance (Geer, 2016); large "step changes" in medium-range forecast quality that arise just as a result of observation changes are increasingly rare. Given this context, the new results are very encouraging and they highlight the benefits of filling observation gaps. Spread improvements are apparent in the globally averaged profiles to around day 3; these are primarily driven by the large improvements seen in the tropics, which are clear at both day 3 and day 5. The agreement between the EDA and ENS results at these forecast ranges is reasonable and within the EDA error bars, particularly in the tropics where the signals are strongest (the ENS results suggest a slightly larger impact at day 5 than the EDA, but in general the EDA and ENS results provide a consistent picture). SC 3 improvements are generally larger than SC 1 when assessed over large areas. The day 5 percentage spread reductions are smaller than those at day 3. For example, the 700 hPa spread reductions in the tropics approximately halve from day 3 to day 5, but they remain statistically significant. This reduction in percentage difference occurs because the absolute spread differences remain similar throughout the forecast range, while the spread value used to normalise the difference increases.

More generally, ECMWF has assessed the ENS results using a "scorecard", which is a web-based verification tool that visualises a comprehensive set of surface and upper-air forecast impact scores out to day 15. While it is difficult to show the full scorecards generated in this study, screenshots and an explanation of key hemispheric scores are given in Annex 1. In general – but noting some variations across pressure levels and parameters – the SC 3 ENS experiments show statistically significant spread improvements in both surface and upper-air parameters to around days 5-6 for the northern hemisphere extra-tropics, and to day 4-5 for the southern hemisphere extra-tropics. The SC 1 impact is generally about 1 day weaker than the SC 3 impact, extending to around day 4 in the northern hemisphere and day 3 in the southern hemisphere.

The largest impact for both SC 1 and SC3 is in the tropics, where statistically significant spread improvements in wind and temperature are typically found out to day 7 for both scenarios, from the surface to 100 hPa (~16 km).

4.2 Impacts by region (localized impacts)

Large surface and upper-air forecast impact for Africa extending to day 6. These improvements spread westward into the Atlantic

Figure 4 shows EDA spread changes for temperature in the smaller, bounded regions illustrated in both Figure 1 and Figure 2, encompassing Africa, western Sahel, a region in the Pacific and Europe. In general, the spread reductions are larger for SC 3 than SC 1

when the impact is large. There are some counter examples (Europe day 5), but this can be attributed to sampling error, noting that the error bars for scenarios usually overlap when the expected ordering appears to reverse.

The SC 3 improvements are statistically significant out to day 5 in the large region encompassing Africa and the smaller area within it, covering the western Sahel. In fact, scorecards for Africa (not shown) indicate that the impact extends beyond day 5. The positive impact of the new observations in the Pacific region extends to day 5 in the lower troposphere, although it is a factor of two smaller than in Africa, where the density of new observations is much higher. Similar results are obtained for profiles of wind and relative humidity.

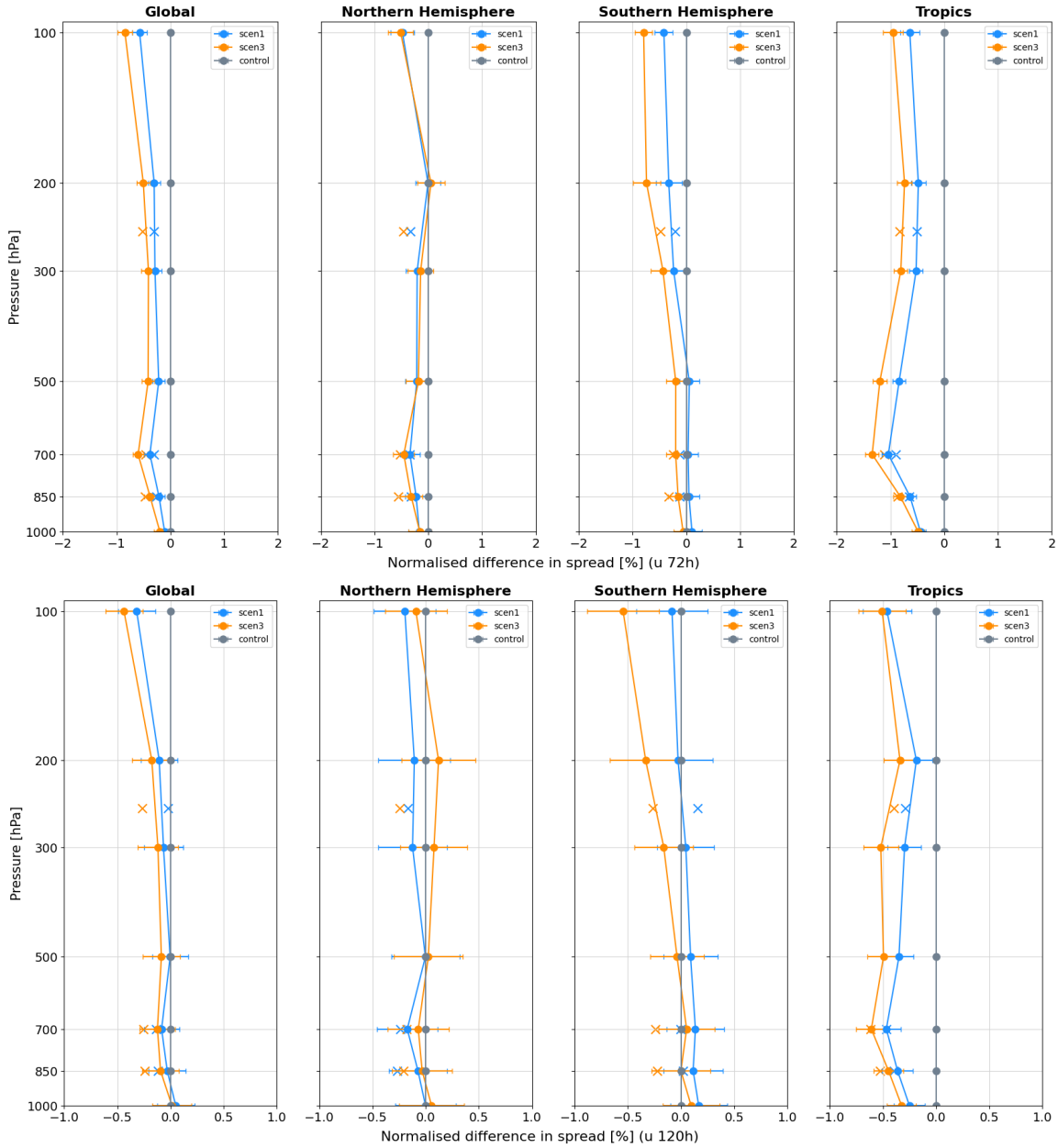


Figure 3. Profiles of normalised spread reductions for zonal wind at day 3 (upper) and day 5 (lower) for SC 1 and SC 3. The solid lines/circles are derived from the EDA, and the crosses at 850,

700 and 250 hPa are spread changes from the ENS experiments. Negative values indicate that the forecast error statistics are improved.

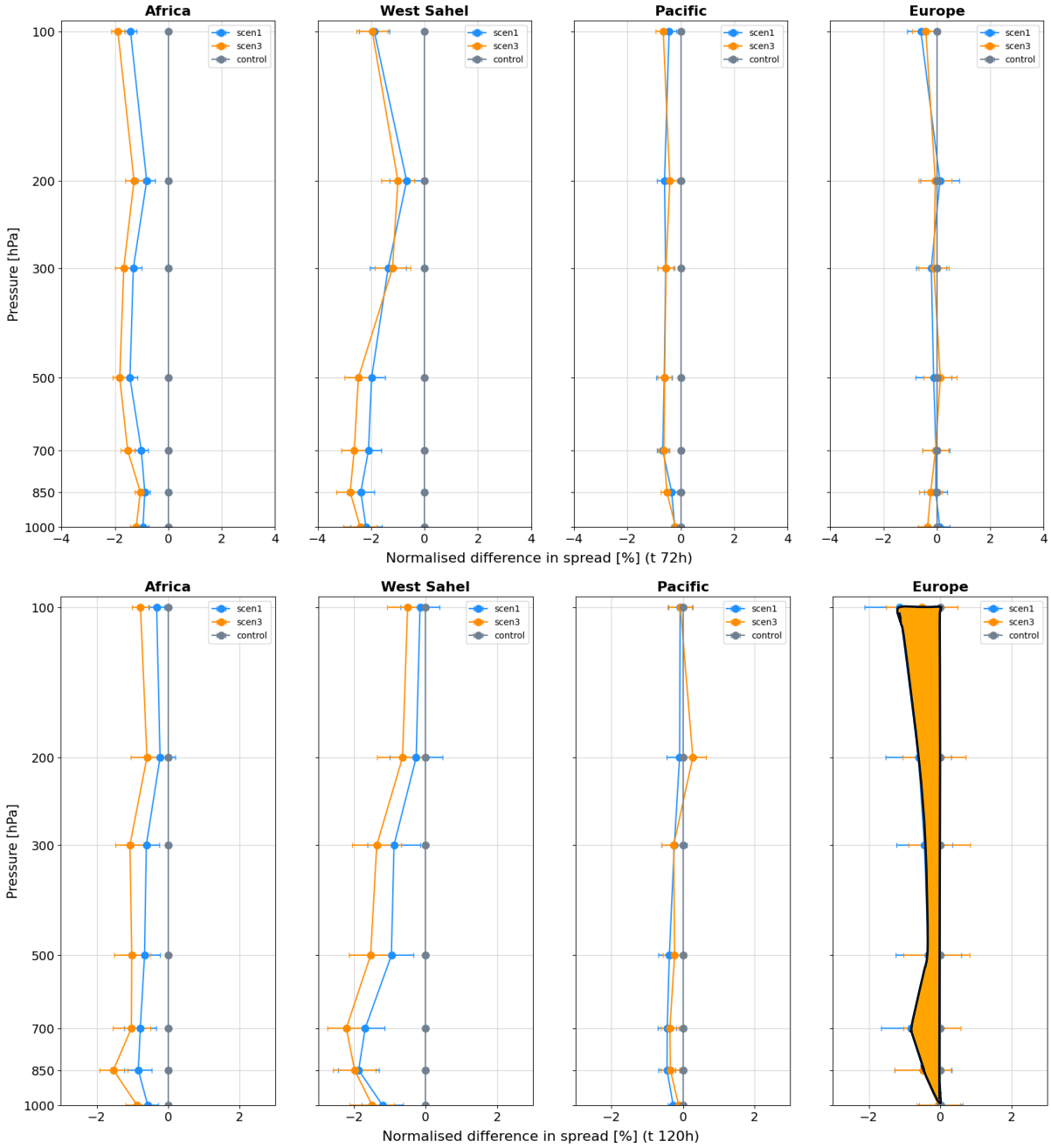


Figure 4. Profiles of normalised spread reductions from the EDA for temperature at day 3 (upper) and day 5 (lower) for both SC 1 and SC 3. Negative values indicate that the forecast error statistics are improved.

Figure 5 shows the absolute spread difference of total column water for SC 3. At the analysis time (top panel) the largest differences are on the African continent and over Central America, where most of the new observations are introduced. The differences are very small in the central tropical Atlantic. At day 1 (mid panel), the differences start increasing in the Eastern Central Atlantic. By day 3 (bottom panel), the differences in the

Central Atlantic have the same magnitude as those in Africa. Overall, the signal (spread differences) is moving west from Africa towards Central America with time. Note that the relatively small regions of increased spread in the North Atlantic shown in Figure 5 can be attributed to sampling error, and they would be expected to be smoothed out with larger ensembles and/or longer experimental periods.

In summary, there is clear evidence that the impact of the new observations spreads beyond the regions where they are introduced as the forecast range increases.

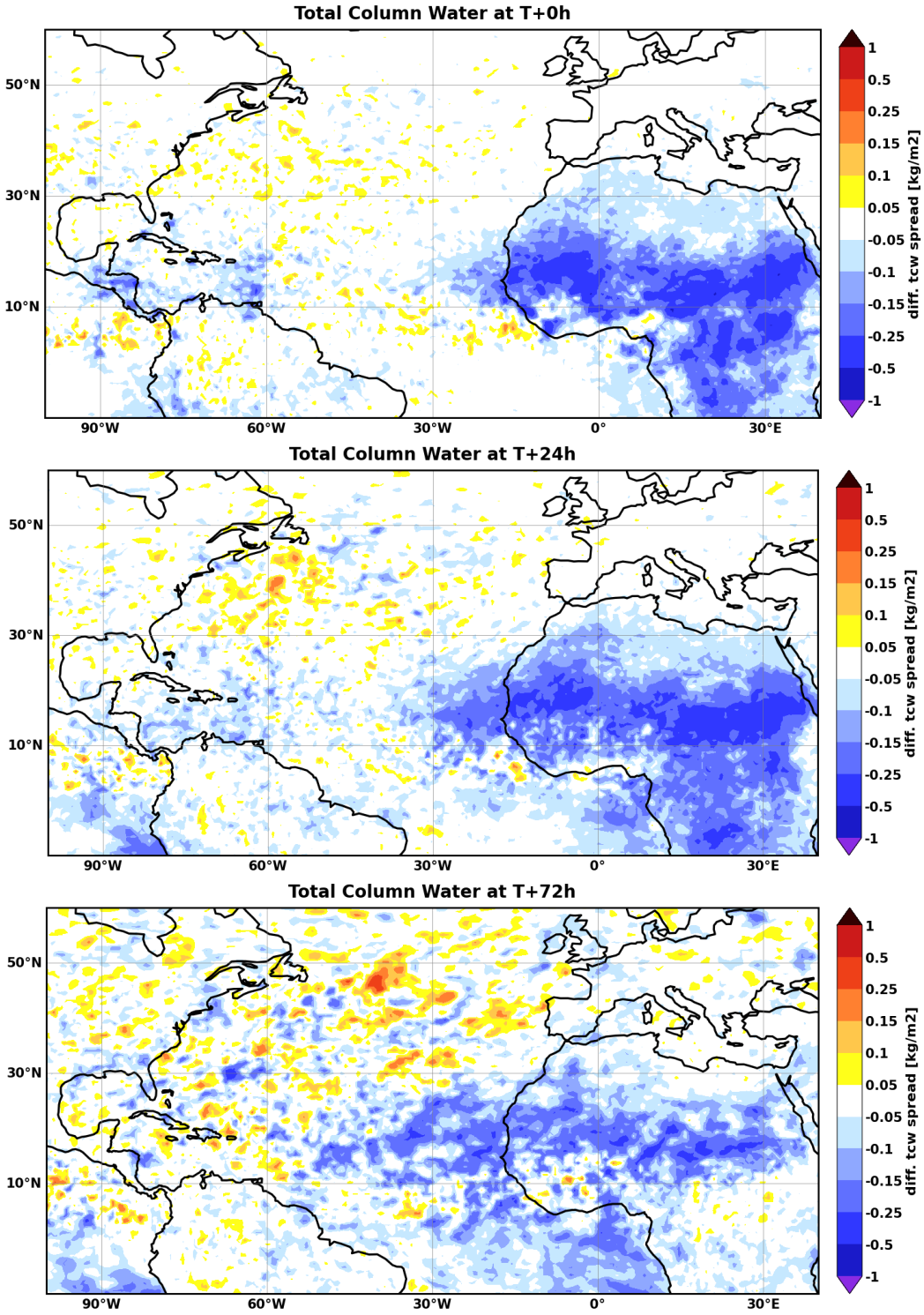


Figure 5. Map of the absolute spread differences from the EDA between SC 3 and the control for the total column water (tcw) at analysis time (top panel), day 1 (mid panel) and day 3 (bottom panel).

4.3 Impact of SOFF expansion on the Atlantic Main Development Region (MDR)

Medium-range forecast improvements in the region where many Atlantic Tropical Cyclones originate

One important result from phase 2 was the reduced initial condition and short-range forecast uncertainty in the Atlantic "MDR" region (defined here as within 6–18°N and 20–60°W), and this is reinforced in the phase 3 results for the medium-range.

African Easterly Waves (AEWs) are usually initiated over northern-central Africa – where the observing network is being improved in both SC 1 and SC 3 – and propagate westward in the lower to mid-troposphere. They typically occur from late spring through autumn, and they play a key role in organizing convection over West Africa. They are the primary precursors for many Atlantic tropical cyclones that form in the MDR between Cabo Verde and the Caribbean. In fact, around 60 % of Atlantic tropical cyclones develop from AEWs, so any potential forecast improvements are important for tropical cyclone preparedness for Caribbean SIDS.

As AEWs have a mid-tropospheric alternating north–south flow, the meridional wind, *v*, at 700 hPa is considered a useful diagnostic. In both the SC 1 and SC 3 EDA experiments, the increased number of wind observations in the lower and middle troposphere over central and western Africa reduces the theoretical uncertainty in the initial conditions. It is found that the improvements in forecast uncertainty extend into the medium range. Figure 6 shows the variation of the EDA ensemble spread of the v wind at 700 hPa with forecast-range for the control, SC 1, and SC 3 experiments, averaged over the MDR region. Both scenarios improve the forecast lead time by around 4-5 hours out to day 5. Similar results are found for other key parameters, for example the specific humidity (Q) at 700 hPa.

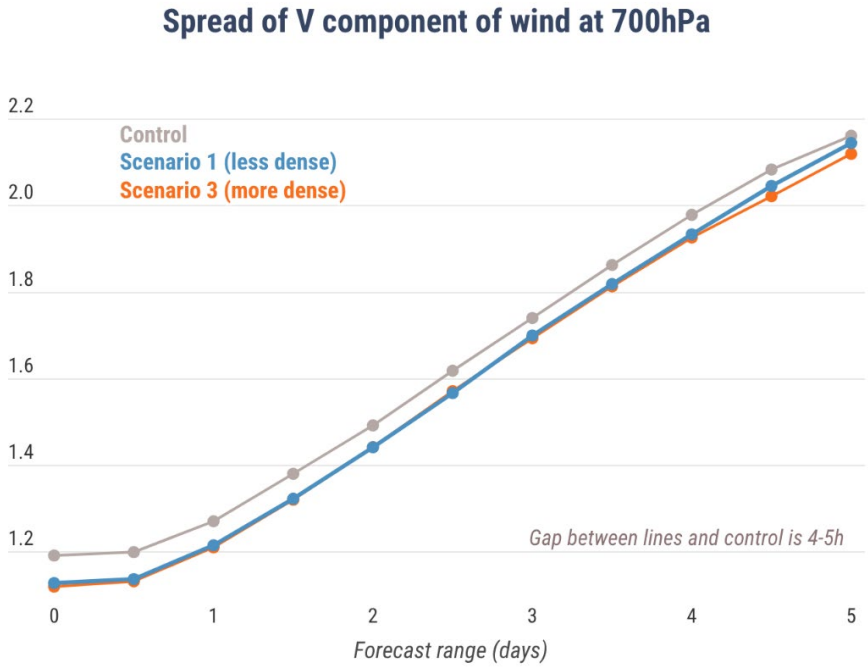


Figure 6. The standard deviation of the ensemble spread as a function of forecast-range for the v wind component (m/s) at 700 hPa in the north-Atlantic MDR. These are computed from the EDA experiments.

Figure 7 shows the time series of the v 700 hPa EDA spread values in the MDR. It is notable that the peak spread value on August 15, 2023 is associated with an AEW that developed into TC Franklin, and spread reductions of around 0.8 m/s for SC 3 are found for that case. However, more generally, it must be acknowledged that the uncertainty in North Atlantic TC track and intensity forecasts was not improved during this five-month period, which still represents a relatively small statistical sample. The spread values of both parameters were computed with a standard ECMWF verification package, but no statistically significant results were found for this 5 month period.

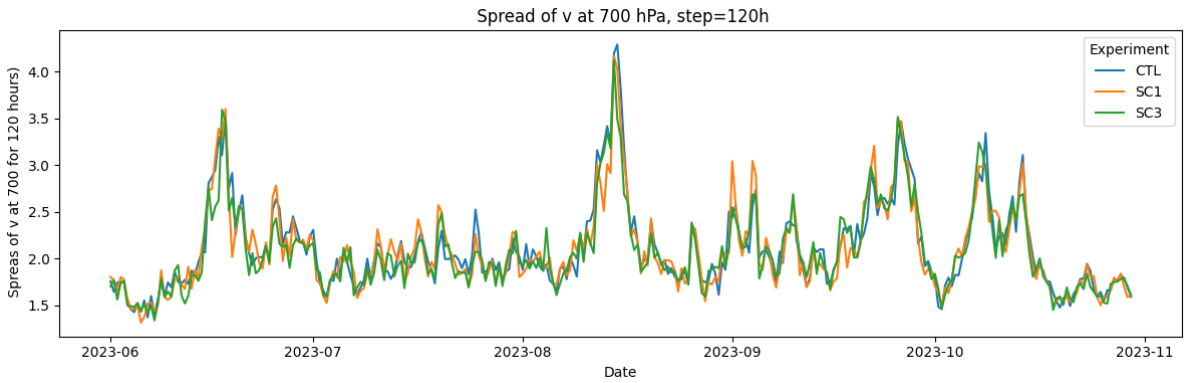


Figure 7. Timeseries of the day 5 forecast v wind component spread values at 700 hPa for the MDR region. These are computed from the EDA experiments.

5. Concluding remarks

The Phase 3 ECMWF impact studies provides further evidence that SOFF-supported observation enhancements in data-sparse regions would deliver significant benefits to global weather forecasting. The new results suggest that these improvements are likely to extend into the medium range, typically out to day 7 in the tropics, with measurable reductions in forecast uncertainty over Africa, the Atlantic, and the Pacific. The potential 4–5 hour forecast lead time gains for monitoring conditions linked to North Atlantic tropical cyclone development is a potentially significant result for early warning and preparedness for the Carribbean SIDS.

6. Next steps

New observing system simulation experiments (OSSEs) would be useful to confirm the key EDA/ENS results contained in this report, including the very promising MDR impacts. In addition, the OSSE technique is probably more appropriate for individual case studies for high impact events if they are required in the future.

Further development of the combined EDA/ENS for medium-range assessment in longer studies is advisable. The implications of the new machine learning data assimilation approaches for improving the exploitation of the SOFF observations should also be investigated.

7. References

Geer, A. J. (2016). Significance of changes in medium-range forecast scores. Tellus A: Dynamic Meteorology and Oceanography, 68(1). https://doi.org/10.3402/tellusa.v68.30229

Healy, S. B., Bormann, N., Geer, A. J., Hólm, E., Ingleby, B., Lean, K., Lonitz, K., and Lupu, C. (2024). Methods for assessing the impact of current and future components of the global observing system, ECMWF Technical Memorandum 916, https://www.ecmwf.int/en/elibrary/81556-methods-assessing-impact-current-and-future-components-global-observing-system

Lean, K., Bormann, N., Healy, S., English, S., Schüttemeyer, D. & Drusch, M. (2025) Assessing forecast benefits of future constellations of microwave sounders on small satellites using an ensemble of data assimilations. *Quarterly Journal of the Royal Meteorological Society*, e4939.

Report from Phase 2 (2025) https://www.un-soff.org/wp-content/uploads/2025/06/ECMWF-SOFF-Impact-Experiments-June-2025.pdf

Annex I

"Scorecards" are a web-based tool used at ECMWF to provide a visual overview of forecast impact for both deterministic and ensemble experiments. They have been used in this study to assess the ENS results which extend to day 15.

Figure A 1 and Figure A 2 show screenshots of the scorecards for ensemble spread changes in the SC 1 and SC 3 experiments, respectively, when compared with the control experiment. These show the spread changes for geopotential height (z), temperature (t), and wind speed (ff) on the 850, 500, 250, 100 and 50 hPa pressure levels, and relative humidity (r) on the 700 and 200 hPa levels. The surface variables, pressure at mean sea level (msl) and 2 m temperature (2t), are also shown. Each box in the horizontal direction represents a forecast day, ranging from day 1 to day 15. The statistics are split into the northern hemisphere extra-tropics (latitude > 20°N), tropics (20°S ≤ latitude ≤ 20°N) and southern hemisphere extra-tropics (latitude < 20°S), shown in the scorecards as "n.hem", "tropics" and "s.hem", respectively.

A green box represents a spread reduction – or improved error bar for the forecast – and conversely a purple box shows that the spread has increased. A bold green or purple box suggests that the change is statistically significant at the 95% level. For example, consider the "2t" row in the "n.hem" columns shown in Figure A 1. The bold green boxes extend to day 4, and this row is interpreted as a statistically significant, positive impact to day 4, and then neutral impact from day 5 onwards. More generally, Figure A 1 shows that the spread improvements in the northern hemisphere are generally positive to day 4 for most variables. The impact in the southern hemisphere is clearly smaller out to around day 3, and the largest impacts are in the tropics, extending to around day 7. Some variations depending on variable and pressure level are apparent (e.g., wind speed in the tropics extends to day 8 at 850 hPa and day 6 at 250 hPa), but the overall ranking of impact – tropics, followed by northern hemisphere, then southern hemisphere – is clear.

A comparison of Figure A 1 with Figure A 2 illustrates that the SC 3 impact is generally larger than SC 1 for all regions. In the northern hemisphere the impact typically extends to day 5 or day 6. In the southern hemisphere, it usually extends to day 4 (e.g., 850 hPa ff) or day 5 (e.g., 700 hPa r). The largest surface and upper-air impact is again in the tropics. For example, the 2 metre temperature (2t) impact extends to day 6, and the 850 hPa temperature and wind speed improvements extend to day 7. The 250 hPa wind speed appears to improve beyond day 10 highlighting the potential value of new wind observations in the tropics.

In summary, the scorecards illustrate that the impact for SC 1 and SC 3 is expected to extend into the medium-range. SC 3 has a bigger impact than SC 1 and the largest impacts are expected in the tropics.

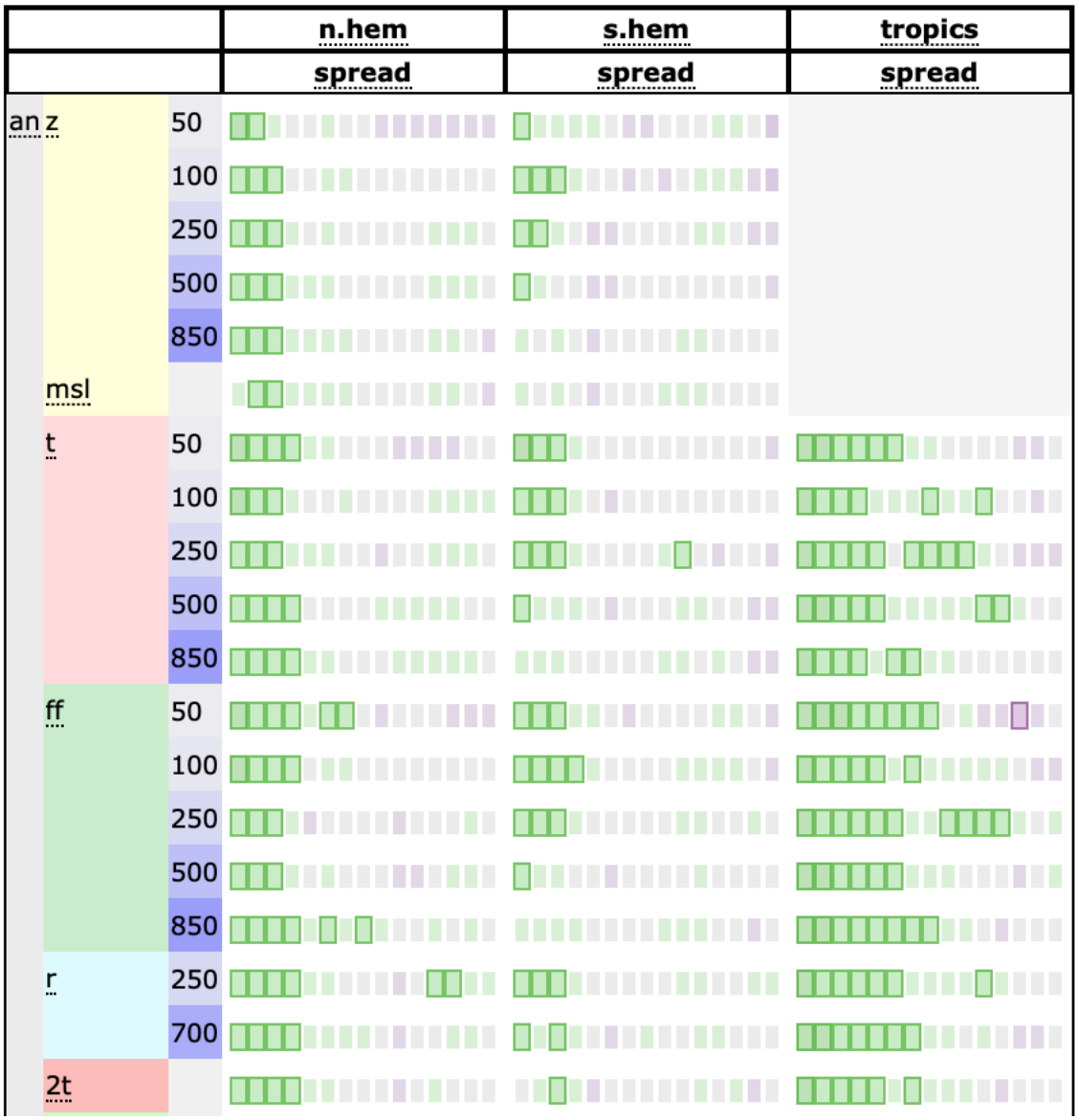


Figure A 1. A reduced scorecard for the SC1 (LDCs+SIDS) ENS experiments. In the horizontal direction each box represents a forecast day, starting from day 1 out to day 15. Green indicates a spread improvement and purple a degradation. A bold green box indicates that the spread improvement is statistically significant at the 95 % level. The statistics are split into the northern hemisphere and southern hemisphere extra-tropics and the tropics. The meteorological variables and pressure levels are shown in the vertical direction. The variables given on pressure levels (z, t, ff, r) are (geopotential, temperature, wind speed, relative humidity). The surface variables (msl, 2t) are mean sea-level pressure and 2 metre temperature.

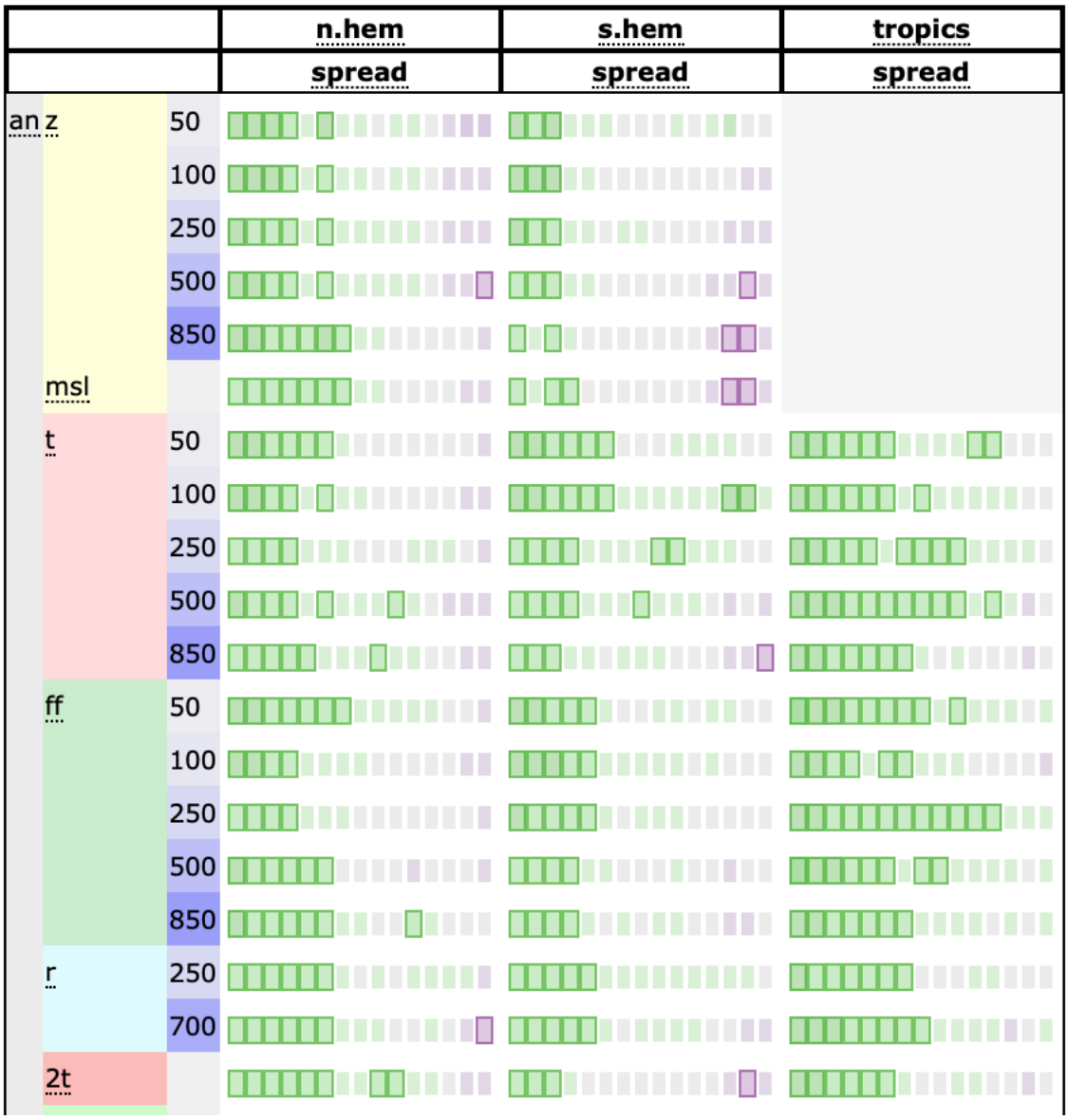


Figure A 2. As Figure A 1, but for SC 3 . (All ODA+SIDS)