Systematic Observations Financing Facility

Ninth Steering Committee
22 October 2024

SOFF impact report: Literature review and proposed SOFF-targeted studies

Decision 9.2

Systematic Observations
Financing Facility

Weather
and climate
data for
resilience

Decision 9.2: SOFF impact report: Literature review and proposed SOFF-targeted studies

The SOFF Steering Committee

Welcomes the first SOFF impact report that reviews the finding of existing impact studies from systematic observations based on this review proposes additional tailored experiments to assess the impacts of SOFF current and proposed investments.

Thanks the European Centre for Medium-Range Weather Forecasts (ECMWF) for conducting the review of existing studies and undertaking the proposed additional studies in collaboration with WMO.

Notes

- that investing in closing the surface-based observations gap is highly beneficial and a key element necessary to improve forecast accuracy.
- that stations in data sparce regions have a much larger individual contribution to the forecast accuracy than any one station in a well observed region, underlying the importance of global coverage by the Global Observing System.
- that existing research and studies provide general evidence for the importance of investments in surface-based observations, but it is recommended to undertake additional SOFF-tailored scientific studies to effectively guide SOFF investment priority decisions.

Adopts the first SOFF impact report and the proposed scenarios for SOFF-tailored further research and experiments to further develop understanding of different impact scenarios to strengthen the evidence base of SOFF investment decisions.

Requests

- ECMWF in collaboration with WMO to conduct these SOFF-tailored experiments to present the results to the 11th SOFF Steering Committee.
- The SOFF Secretariat to share the SOFF impact report with the WMO Commission for Observation, Infrastructure and Information Systems

Purpose of this Document

This document, prepared by the European Centre for Medium-Range Weather Forecasts (ECMWF) in collaboration with WMO, summarizes existing scientific studies on the impact of observations in forecast skill.

It concludes that surface-based observations significantly impact forecast accuracy and that there is substantial evidence that investing in additional surface-based observations is highly beneficial.

It further concludes that more scientific studies on the impact of surface-based observations are required to effectively guide SOFF investment priority decisions. It proposes scenarios for SOFF-tailored further research and experiments to fill existing knowledge gaps.

Table of contents

Executive Summary

SOFF Steering Committee decision to produce SOFF impact reports

The SOFF Steering Committee, through Decision 6.8, decided to produce initial SOFF Impact Reports as part of the SOFF Work Programme 2022-2025. Impact is defined as the impact of improved observations in forecast skill.

The reports are being conducted in two phases. The first phase focuses on a summary of the findings of existing scientific studies, i.e. answers the question "what do we know and what do we not know". It also proposes scenarios for SOFF-tailored further research and experiments in phase two in order to further develop understanding of the impact specifically for SOFF investments. This document is the output of the first phase. The results of the second phase will be presented to the 11th SOFF Steering Committee, expected to take place in May 2025.

WMO has engaged the European Centre for Medium-Range Weather Forecasts (ECMWF), one of the World Meteorological Centres and a member of the SOFF Advisory Board, to undertake the studies.

Observations are the foundation for all weather forecasts

All monitoring and prediction of weather start from the collection and global exchange of observations – these data provide the only direct source of information about the atmosphere. Weather is inherently global, and to understand and predict them anywhere, observations covering the entire globe need to be made available to the global monitoring and prediction model systems.

Spaceborne and surface-based observations

The WMO Integrated Global Observing System comprises both, satellite- and surface-based observations. Spaceborne observations offer global coverage, capturing data from every part of the Earth. However, the data they provide is often complex and requires interpretation to extract useful information. 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. However, there are currently large data gaps on surface-based weather observations, negatively affecting the quality of weather forecasts everywhere. Closing these data gaps is essential for the world to be better prepared, better understand and therefore to effectively adapt to a changing climate.

World Meteorological Congress break-through decisions

In 2021, the 193 countries and territories of the World Meteorological Congress took three inter-related break-through decisions. For the first time it defined what constitutes the minimum set of surface-based weather observation that all countries are mandated to generate and internationally exchange - the Global Basic Observing Network (GBON). It also decided that these data have to be freely and openly available. Finally, it decided to establish the Systematic Observations Financing Facility (SOFF) as a new UN climate fund to support countries in achieving sustained GBON compliance. SOFF investment decisions need to be based on evidence where to invest to maximise benefits, emphasising the importance of the SOFF impact reports.

Main findings from existing scientific studies

Existing studies show that surface-based observations significantly impact forecast accuracy, especially in areas with sparse data coverage.

- High impact of surface-based observations: Despite being fewer in number compared to satellite data, surface-based observations have a strong influence on forecast accuracy in many regions. It is also known that, in general, adding one new surface-based observation in a data sparse region has more impact than adding a similar new surface-based observation in a data-rich region.
- 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.
- Importance of both surface land and upper air stations: while both are critical elements of the global observing system, the relative importance of surface land versus upper-air observations varies across studies. More studies suggest that upper air observations have more impact, and there is a tentative acceptance that upper air observations have more impact than near-surface observations.
- Significant geographical observation gaps: There is evidence that data gaps in Africa (in particular regions such as East Africa, the Rift Valley, and the Horn of Africa), parts of the Pacific and Atlantic Oceans, the Arctic Ocean and Antarctica, need to be reduced for improved forecast accuracy.
- More scientific studies on the impact of surface-based observations are required. Most studies in the literature explore satellite data impact, studies on surface-based data are more limited. Existing studies on the impact of existing observations do not provide the level of granularity required to fully guide SOFF investment priority decisions.

In summary, there is substantial evidence that investing in additional surface-based observations is highly beneficial.

However, the existing literature gives only limited guidance for SOFF prioritization decisions, which demonstrates the need for the additional SOFF-tailored studies. These studies will employ techniques previously used to assess the impact of satellite data, but they will focus on surface-based observations. By simulating the addition of new observations through experiments, these studies aim to provide clearer insights into the identification of investments that would yield the most significant benefits.

Proposed scenarios for SOFF-tailored Observation System Experiments

To assess the potential impact of additional surface-based observations, several scenarios are proposed:

- Baseline: Current observing system.
- Ensemble of Data Assimilations Scenarios: Adding simulated surface land and upper-air observations for different country groupings (Small Island Developing States [SIDS], Least Developed Countries [LDCs], Lower-Middle Income countries, all Overseas Development Aid [ODA]-eligible countries, Fragile and Conflict Affected States) and regions (Africa, Pacific).
- Observation System Experiments Scenarios: Simulating the absence of surface land and upper-air observations.

These scenarios will be tested to determine how additional observations would improve weather forecasting especially in areas currently lacking sufficient observations.

1. Review of existing studies

Since the start of the satellite era, global observations have been available, yet the impact of surface-based observations remains highly significant. For example, Pauley and Ingleby (2022) noted that despite only accounting for 11% of observations used, radiosonde, aircraft and other surface-based observations together accounted for 26% of the forecast error reduction of a day one forecast at ECMWF, and 29% for the weather forecast system operated by the United State Navy's Fleet Numerical Meteorology and Oceanography Center, in January 2020. This reminds us that, even in the satellite era, surface-based observations continue to have significant impact. The availability of surface-based observations is not uniform globally, so neither is their impact. Their availability and hence total impact is low in many parts of the southern hemisphere, though the value of individual observations is higher in data sparse regions (Ingleby 2021). Even in the northern hemisphere there are locations where coverage falls well short of the Global Basic Observation Network (GBON) requirement. The GBON concept was adopted by the 2021 Extraordinary World Meteorological Congress of Technical Regulations for the Global Basic Observing Network (GBON) per Resolution 2 (Cg-Ext(2021). GBON defines spatial and temporal requirements for surface-based data provision. Given the need to prioritize SOFF investments to support countries in achieving sustained GBON compliance, it is useful to provide guidance on where the highest impact from adding new observations may be found.

In this report we briefly review relevant existing studies. These studies employ existing well-established techniques that are described in the next section. We find that these studies do not provide all the answers needed. In the final section, scenarios for new dedicated experiments using ECMWF's Ensemble of Data Assimilations (EDA) will be listed. This EDA approach has been used for over a decade to guide investment in spaceborne observations (e.g. Harnisch *et al*. 2013), but this will be the first time it has been applied to surface-based observations.

1.1. The impact of *existing* observations

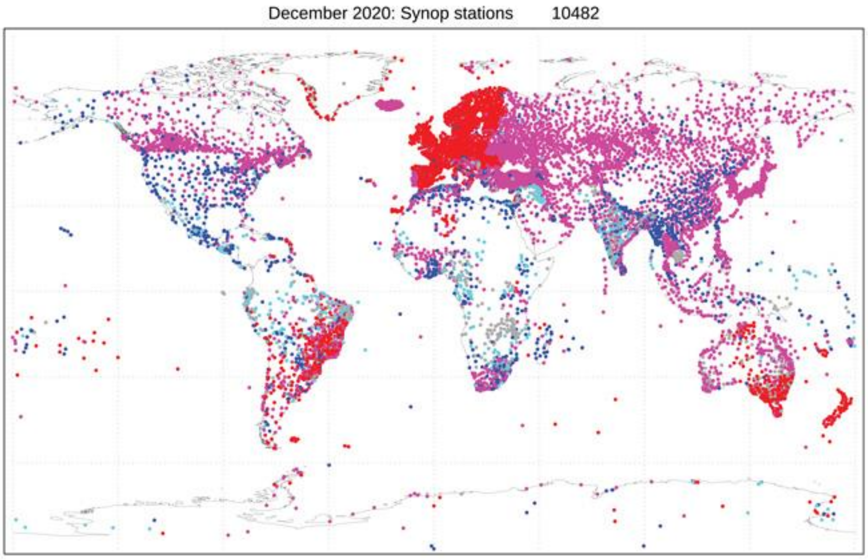

Figure 1. Active land surface stations (SYNOPs), December 2020. The colour gives the frequency of reports received: grey - fewer than 2/day through to red - 24/day. The WMO Integrated Global Observing System (WIGOS) is now encouraging the global exchange of hourly reports.

At the 46th meeting of the Coordinating Group for meteorological satellites (CGMS) a position paper (English, 2018) was presented that described the strengths and weaknesses of the two main methods used for assessing the impact of existing observations, OSEs (Observation System Experiments) and FSOI (Forecast Sensitivity Observation Impact). In brief, and without going into technical details, English (2018) noted that OSEs answer the question "what happens to forecast skill if I lose or add this data type?" FSOI answers the question "in the operational configuration how much does this data type contribute to short range forecast error reduction?". There is a similar technique to FSOI called Ensemble-based Forecast Sensitivity to Observations (EFSO) which for the purposes of this paper can be considered equivalent to FSOI. In the following discussion key relevant messages from published studies using these approaches will be summarised.

It is important to note methods such as OSE and FSOI can only tell us the impact of observations used by or available to be used by the forecast system. In this sense they cannot tell us directly the impact of additional observations not yet available, but that might become available in the future. We can in Figure 1, from Ingleby (2021), see where the observation gaps occur. Notably several parts of Africa, Tibet, the oceans and the polar regions stand out. It is therefore of interest to discuss impact of existing observations, but important to note this cannot fully answer the question on impact of new observations.

1.2. Literature review of relevant observation impact studies

1.2.1 Observation System Experiments

Observation System Experiments (OSEs) are the gold standard for assessing the impact of observations. However, they do have weaknesses. OSEs compare the forecast skill of an experiment that adds or removes an observation type to the skill of a control experiment. The observation impact found will depend on the quality of the control. However, there are many flavours of OSE, which means results need to be interpreted carefully. For example, sometimes other observations are removed to create a baseline which has a much lower accuracy than normal. It is easier for new observations to show an impact when assimilated in such a system, so the size of their impact is amplified. This is useful when checking technical functionality of the DA system but is not a useful indicator of the impact of the observations in current operational systems. There is also a flavour of this where the background in the data assimilation comes from an analysis that had seen all observations, but the new analysis for the forecast uses a data depleted control. In both types these OSEs do not measure the observation impact in the current state of the art system and have little relevance to decision making for future investment. Therefore, in this report we will give more weight to OSEs for which the impact is assessed using data removal against a state-of-the-art observation-rich control.

Unfortunately, OSE studies are often published for very short periods, perhaps focused on individual storms, so care is needed to check the statistical relevance of results. Short experiments can produce eye-catching results, and can be important to illustrate impact, especially if the forecasts with and without the observations are very different, and the storm had a high impact. However, such results have no statistical significance, and it is important not to be selective to pick cases with large beneficial impact, and neglect cases with no or even negative impact. This does not mean case studies and short experiments have no value, but they do not answer the question we are addressing in this report. In this report we will base our conclusions primarily on experiments whose statistical significance is demonstrated.

1.2.2 Reports from observation impact workshops and conferences

Results from OSEs have been reported at major fora that attempt to summarise the current value and impact of the Global Observing System.

WMO organises regular “OSE workshops”, the first was held in Geneva in 1997 and the most recent 8th workshop was held on May 27-30, 2024 in Sweden (WMO 2024). These workshops provide a high-level summary of the impact of the main observing systems, and the reports show how the impact of different observations has evolved over the years. Although the impact of satellite data has grown, these workshops continue to show that surface-based observations retain importance. The reports are summarised e.g. from the 8th WMO OSE workshop at WMO (2020, 2024). Furthermore, key findings of a joint symposium between WMO and ECMWF are described by Valmassoi et al. (2023).

The report from the 8th WMO OSE workshop (WMO, 2024) noted significant differences between different forecasting organisations on the relative importance of different observing systems, in both regional and global NWP systems. This was also noted in Valmassoi et al. (2023). In addition, the application of Artificial Intelligence to weather forecasting is now demonstrating competitive forecast skill (Ben-Bouallegue et al. 2023; Kitchen, Bennett and Chantry, 2023) and at the present time it is still not certain how much the use of this new machine learning technology will have on the requirement for new observations. This means there is no absolute answer as to the value of an individual observation. This depends on differences in the scientific and technical maturity of different systems, the spatial and temporal resolution, global or regional, and the forecast range of interest.

Despite the different conclusions emerging from individual studies, WMO (2020, 2024) did provide some overarching messages relevant to this study. For example, it has been found to be beneficial to use more upper-air surface-based observations from weather balloons. In particular results showed benefit if more data is collected as the balloon rises (in the past, due to operational constraints, only a limited number of observations were reported as the balloon went up) and also to show impact as the sensor falls back to earth after the balloon burst, which in the past was not collected. Benefits were also reported widely from assimilating aircraft observations at higher spatial resolution and, for the first time, from aircraft humidity. It was also reported that additional radiosondes provided for the Year of Polar Prediction improved forecasts. Whilst these results are not directly relevant to the deployment of additional observation sites, they do underline that adding more data from surface-based observation systems is generally found to be beneficial.

Similar conclusions were reached from a number of presentations at the ECMWF Annual Seminar on observations, summarised in Valmassoi et al. (2021). These results confirm we continue to see impact on forecast quality from a wide range of additional surface-based observations.

1.2.3 Reports from observation impact publications

In addition to the events aiming to capture a consensus, many individual studies are published on impact of observations (e.g. since 2017: James and Benjamin 2017; Ingleby and Isaksen 2018; Ito *et al.* 2018; Schäfler *et al.* 2018; Bormann *et al*. 2019; Lawrence *et al*. 2019; McNicholas and Mass 2021; Yamazaki *et al*. 2021; Ingleby 2021; ECMWF 2021; Randriamampianina *et al.* 2021; Liu *et al*. 2022; Laroche and Poan 2022; Demortier 2023; Chambon *et al*. 2023). Whilst there is a large literature on OSEs, these are usually referenced against a baseline which includes all surface-based observations and consequently OSEs usually measure impact of adding different types of satellite observations. There are surprisingly few studies studying the value of surface-based observations on a global scale. An exception is the recent study by Chambon *et al.* (2023). In general, these papers show the overall impact of major observing systems, and highlight the level of agreement and consistency, but give little more than hints where investment should be made. The most comprehensive sets of long-period OSEs are run by the global NWP centres (Bormann *et al*., 2019, Chambon *et al*. 2023). These show the high value of satellite observations, but also the importance of surface-based observations, both land and marine.

Chambon *et al*. (2023) reported the impact of different observation types in the Météo-France global NWP system. They ran a full observing system experiment as a control, then removed major observation types one at a time, including in one experiment all surface-based observations. They showed that surface-based observations provide by far the largest impact in the northern hemisphere, but this is not the case in the tropics, where the radiances dominate for forecasts of temperature and humidity, and the atmospheric motion vectors dominate for forecasts of wind. The results show that, at least in the northern hemisphere, satellite data is unable to make up for the loss of impact when surface-based observations data are removed.

The Chambon *et al*. (2023) study is also of interest as they went on to explore the relative impact of the upper air surface-based data (radiosonde and aircraft) versus near-surface surface-based observations (synop, ship, buoy). Contrary to studies reported in WMO (2021, 2024) they reported a larger degradation removing surface pressure observations than the upper air data. Whilst their result is striking and surprising, care is needed. This result is found in the Météo-France system and is not replicated elsewhere. We know from existing OSE intercomparisons, e.g. from the WMO workshops noted above, that different centres do find different observation types dominate, depending on details in their individual configurations. We continue to tentatively conclude upper air observations have greater value than surface land observations, but in the light of the Chambon study more work is needed to confirm this. What is clear is that both surface land and upper air surface-based observations continue to have an impact, and additional data increases this impact.

It is also noted that new research concepts may provide estimates of surface pressure from oxygen band radars e.g. Privé *et al*. (2023); Battaglia *et al.* (2023). If these work well, it will change the requirement for surface-based observations of surface pressure. It should also be noted that data assimilation is not conservative, and without mass fields there is a risk of loss of mass during long experiments, which in the past has led to disproportionate, but largely meaningless impacts of surface pressure observations (Dumelow 2004). More recently, Healy (2013) showed that GNSS radio occultation also provides a constraint on the surface pressure, showing an accuracy of around 1 hPa can be achieved without assimilating any surface-based observations only satellite data. As the Météo-France systems uses significant amounts of GNSS observations, it is less likely the Chambon *et al*. results are simply down to correcting large biases. So we can tentatively accept that surface pressure observations at high density are more important than a low density of upper air observations, but this conclusion is one that will need further testing in the next phase of the study. James and Benjamin (2017) found high impact from surface-based upper air (in this case aircraft) observations in a regional local area model even over the North American continent.

Ingleby and Isaksen (2018) found the impact of observations over the Atlantic basin appears larger than over the Pacific. This is particularly true in the storm track region. In the polar regions there is strong impact from surface-based observations, which was observed by Bormann, Lawrence and Farnan (2019) who found, especially in winter, that surface-based observations had very high value in polar regions.

Many studies have shown that information propagates from one region to another e.g. English, Poulsen and Renshaw (1999), Ingleby, Rodwell and Isaksen (2015), Lawrence *et al*. (2019), Yamazaki et al. (2021) and Laroche and Poan (2022) all reported evidence that information from observations, both surface-based observations and those from satellites, could propagate (e.g. through Rossby waves) to produce a smaller but still significant impact on medium range forecast scores in other regions. Therefore, whilst short range impact may be most influenced by adding local observations, all regions may benefit most in the skill of medium range forecasts by addressing observational gaps, even if far from the location of interest. As is noted below, this type of long-range impact is captured well by OSEs but will be missed by techniques based purely on short range impact, as discussed in the next section.

Lastly, in considering data gaps, we should recognise that in the future we may address gaps with unconventional observations, outside the official meteorological networks. The increasing availability of surface-based meteorological observations from non-standard sources (e.g. smartphone air pressure) do provide an opportunity to assess the potential impact of additional surface-based observations (McNicholas and Mass 2021, Demortier 2023). These studies provide a review of studies of various non-standard weather observations, in the context of analysing high spatial resolution features in mesoscale local area NWP models. To date most studies using these unconventional observations have not been performed in the context of global NWP, but a pilot study at ECMWF is now exploring if impact of these observations found in regional forecast systems can be replicated in a global forecast system (Falk et al. 2023). It is difficult to disentangle the issues of assessing how good these new observation types are from the question of gap filling, but the limited studies carried out so far do suggest there is potential benefit to filling known gaps.

In addition to OSEs, sometimes large shifts in the global observing system provide additional insight. The Covid pandemic provided an opportunity to study the impact of prolonged loss of observation reports from aircraft. Ingleby et al. (2021) found no obvious loss of weather forecast skill during the pandemic, despite loss of many observation types, especially in 2020. However, Ingleby et al. (2021) was also found that upper tropospheric forecasts in 2020 would have been even better with a 'normal' level of aircraft data.

1.2.4 Forecast Sensitivity Observation Impact

FSOI (Langland and Baker, 2004) and ensemble-based (EFSO, Kalnay *et al.* 2012) do not require dedicated large experiments but can be computed "on the fly" in an operational environment. They have the great advantage of being able to assess individual observations and integrate over a given selection of observations, for example to integrate the impact of individual stations over a given region. Results also can be averaged over very long time periods removing sensitivity to short period even seasonal variations. It is also produced routinely at many centres, meaning intercomparison between centres is relatively easy. The Joint Centre for Satellite Data Assimilation maintains an intercomparison of FSOI scores at https://www.jcsda.org/jcsda-project-ios for early 2015. At ECMWF FSOI suggests that satellites gave about 76% of total observation impact in 2022 and surface-based observations gave 24%, despite satellites providing 98% of the data received at ECMWF by volume (though after thinning and channel selection this falls to 82%).

Whilst FSOI is clearly useful, English (2018) and Ingleby (2021) note FSOI does have limitations which are discussed in more detailed in Eyre (2021, 2023). Perhaps the most critical is it fails to capture medium range forecast impact. For example, we know poor six-day forecasts over Europe can be traced back to errors originating in the tropical east Pacific or Canada (Magnusson, 2017). This type of impact is missed by FSOI but captured by OSE. Furthermore, FSOI uses a single metric that may neglect moisture or the stratosphere and to create different metrics needs the processing to be rerun. Therefore, great care is needed in interpreting FSOI results. Nonetheless it can give useful insights when interpreted carefully, so a summary of some key studies follows.

1.2.5 Reports from FSOI publications

FSOI has already been used to explore questions of individual radiosonde stations. In Figures 2 and 3, taken from Ingleby (2021), radiosonde (Figure 2) and SYNOP (Figure 3) FSOI is compared between Pacific Island stations where surface-based observations are very sparse, and selected continental areas, where there is a high density of stations. As expected isolated stations, i.e. stations in data sparce regions, have a much larger individual contribution to the FSOI metric than any one station in the well observed regions (reporting frequency, or the vertical resolution of radiosonde reports, also plays a role). This result reconfirms the importance of obtaining observations in less well observed areas, in turn underlying the importance of global coverage by the Global Observing System.

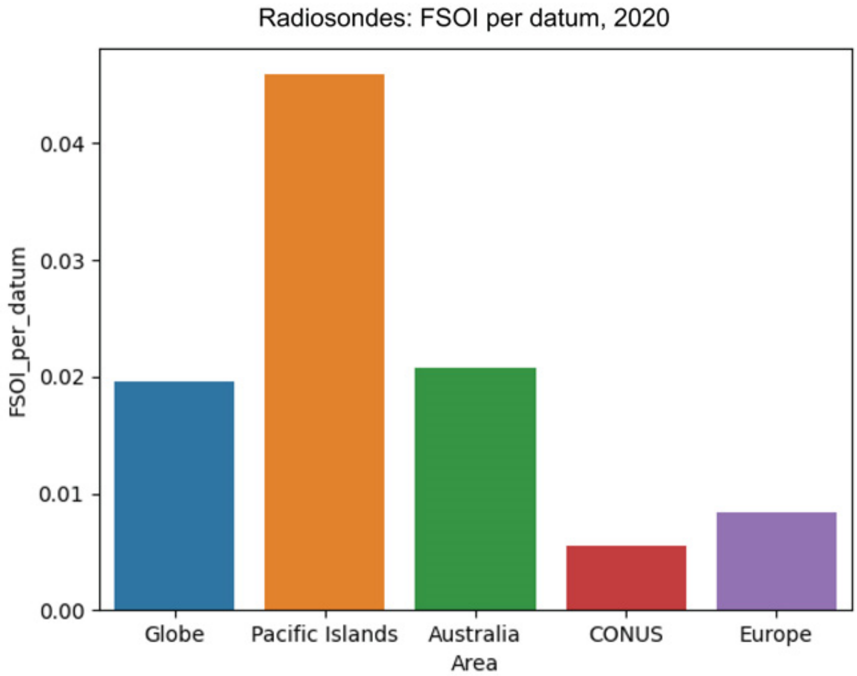

Figure 2. Radiosonde FSOI per datum for 2020 for various regions. CONUS = contiguous USA.

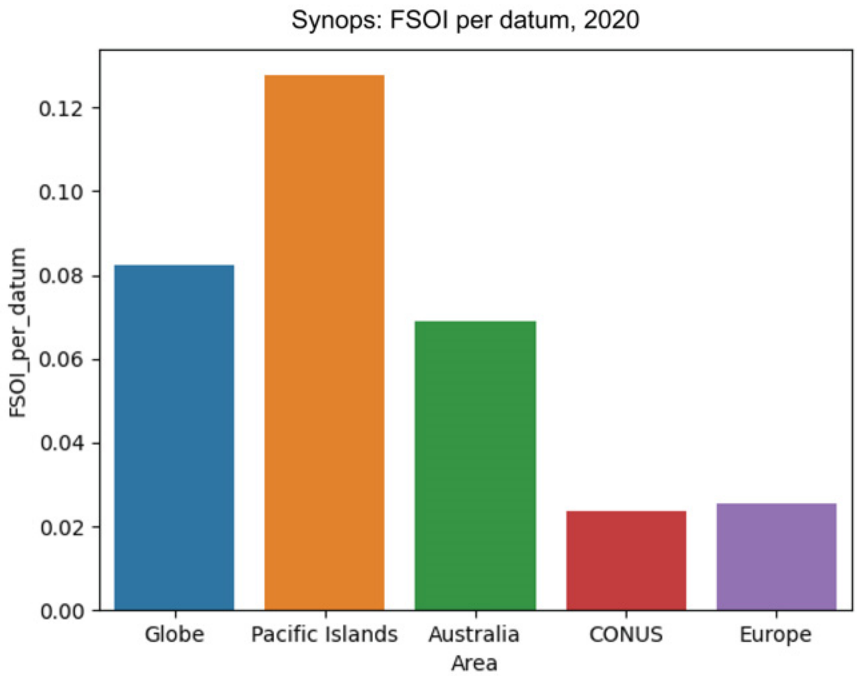

Figure 3. FSOI per datum for SYNOP stations in 2020 - various regions. CONUS = contiguous USA.

FSOI can also give useful pointers to potential areas lacking observations by looking at impact of satellite data. Figure 4 shows a map of FSOI for the IASI sounder on the Metop-C satellite. In general impact is larger over the ocean than the land, as we might expect for satellite observations. But there is large impact over much of Africa, and to a lesser extent south America, Canada and some parts of Asia. A plot like this may give an indication where we are most lacking surface-based observations, by showing where the satellite data is working hardest to fill the gap.

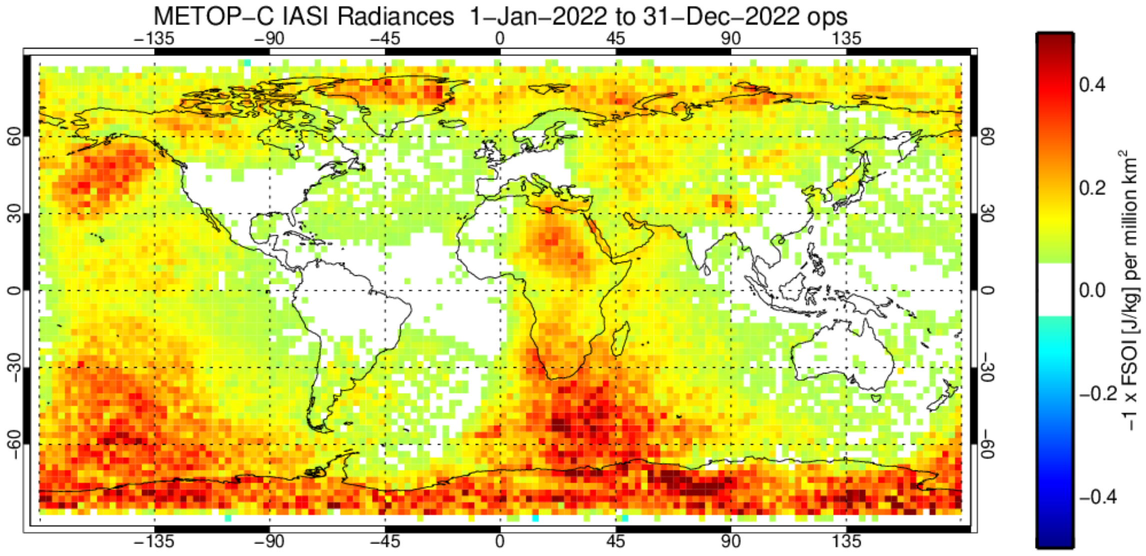

Figure 4. FSOI for the IASI temperature sounder on Metop-C averaged across the whole of 2022.

In Section 1.2.2 OSEs shows a high value for drifting buoy sea level pressure observations, and FSOI also confirms that, per observation, this is one of the most impactful observation types (Centurioni et al., 2017; Horányi et al., 2017).

1.3. Reports on assessment of future observations

In Section 1.2 the impact and value of existing observations was discussed. As was shown there are different approaches but all benefit from a simple fact: you can run the forecast system with and without an existing observation and see how much difference it made. If we want to consider the impact of observations which do not yet exist, the task is more difficult. Therefore, there is a need for quantitative assessment of the impact of potential future deployment of new observations, whether they be ground based or spaceborne, is a more difficult task. There are a range of methods with differing complexity and cost, briefly outlined in the next paragraph. All have pros and cons. All can provide some new insight, and all are an improvement on simply assuming impact always scales with the number of observations and that deployment in all areas is equivalent. Below we outline the method used in this study in context.

At the simplest level we can compute the analysis error covariance and compare to the background error covariance to see the impact of a new observation type. This requires the calculation of a Hessian matrix from the observation and background error covariances, and the gradient of the model relating observations to geophysical space, referred to as the observation operator (Eyre 1990). This Hessian matrix can tell us the theoretical information content of observations on meteorological quantities of interest. We shall refer to this as the Hessian method, which is unique in not requiring the creation of datasets of simulated observations. As noted, this Hessian approach needs something called the background error covariance. This is simply a matrix which tells us how accurately we knew the atmospheric state before observing it. In weather forecast systems forecast errors grow in time (the forecast error for today is usually smaller than for tomorrow, and much smaller than for next week). At very short-range forecast errors are small. This means in a system with lots of observations we have a good estimate before observing again. If we assume background errors are very high, new observations will appear to have high impact on subsequent forecast skill, which they may not have if we assume more realistic background errors. Therefore, for the Hessian approach, and in fact all approaches described here, a realistic background error covariance is essential to gain a realistic estimate of the impact of new observations. It is computed from the operational ensemble of data assimilations (EDA), so we have an estimate which varies with changing weather conditions (some weather regimes have lower forecast error than others).

Simulated observations can be created from a high resolution independent short-range forecast interpolated to a defined time and location, and with random error added consistent with the level of error we expect to find in real observations. The simplest use of synthetic observations is to perform a local one-dimensional variational analysis (1D-Var) assimilating synthetic observations, then computing the analysis departure from our estimate of the truth used to create the synthetic observations (e.g. Deblonde and English 2003). We shall refer to this as the 1D-Var method. This gives a similar 1D answer to the Hessian matrix, but can, to some extent, allow for non-linearity in the observation operator. But it is only a local estimate and lacks any understanding of the multi-dimensional analysis problem that must be solved for operational weather forecasts. Thirdly we can take the 1D-Var concept to another level by assimilating the synthetic observations into an Ensemble of Data Assimilation (EDA), such as ECMWF's ensemble of 4D-Vars, and measure the change in spread, which is taken as a proxy for forecast error. This can be calibrated by running the EDA for real observations in OSE mode, and checking how well the change in EDA spread matches the change in error estimated using real observations. Figure 5 illustrates this comparison for one month period in 2020, for a range of latitude bands and altitudes, for COSMIC-2 radio occultation observations. In general, the predicted spread change with simulated observations matches very well the actual change with real observations. This increases confidence in the EDA method, but still assumes that EDA spread is also a good indicator of forecast error statistics at longer range.

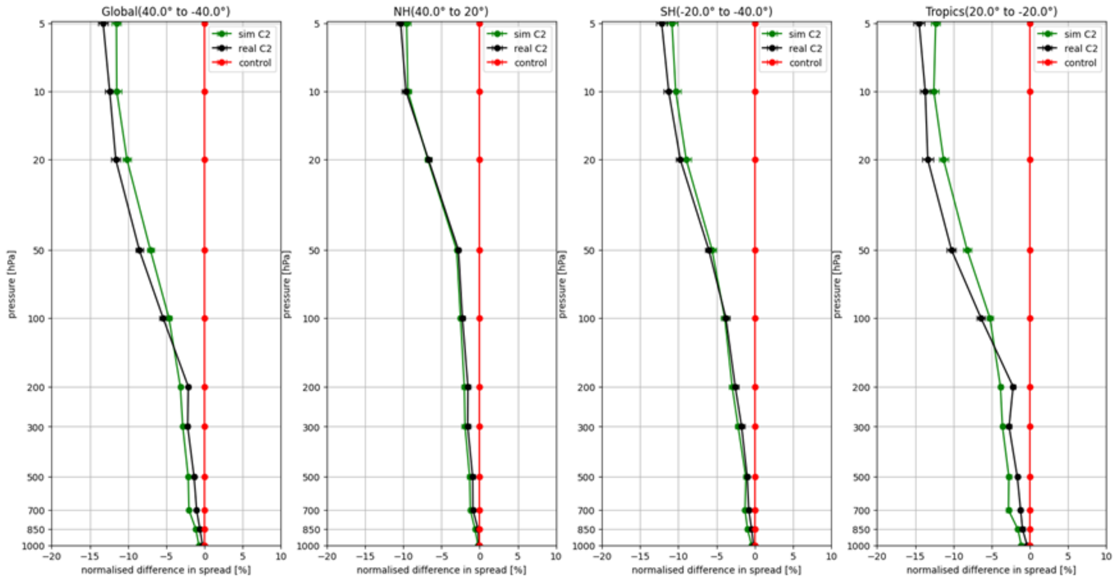

Figure 5. The vertical profile of 12 hour forecast temperature spread reductions, comparing the impact of real (black) and simulated (green) COSMIC-2 measurements in the EDA. The results are given as a percentage of the temperature spread of the control experiment. The spread is computed for the period January 10 to February 10, 2020. The comparisons are limited to ±40 latitude band sampled by the COSMIC-2 data. The control experiment includes GNSS-RO measurements used operationally in this period.

This method allows for 4D aspects to be included, and for cycling through the DA. We shall refer to this as the EDA method (Harnisch *et al*. 2013).

Lastly the most elaborate solution is to simulate all observations and assimilate into an (ideally) independent system to the one that generated the data for the simulated observations, and then run full DA and forecast impact experiments, verifying against the “truth” or nature run used to generate the simulated observations. This is known as an Observation System Simulation Experiment (OSSE) and we shall refer to it as the OSSE method (e.g. Prive et al 2023 for a recent overview). The Hessian is easy to compute, so the cost of such a study is very low, but its applicability is also limited. The 1D-Var and EDA approaches are probably comparable in technical setup costs, though the compute cost for EDA is much higher. An OSSE study is on a different level, with much greater technical investment needed, very high computational costs and a more complicated analysis. Therefore, the EDA approach appears a good compromise between the over simplified analysis from the Hessian or 1D-Var and the expensive solution of an OSSE.

To date most studies that have been performed using Hessian, 1D-Var, EDA or OSSE approaches are for satellite observations. This is because normally they are performed as part of the justification for a large and expensive satellite mission, where quantitative assessment is needed (e.g. Boukabara et al. 2018; Prive et al. 2023). By contrast surface-based observations are funded by National Programmes or more recently by the Systematic Observations Financing Facility (SOFF), as a contribution to WIGOS. The impact of surface-based observations has been assessed using more qualitative arguments.

However, some quantitative studies have been published, mostly to support studies of targeted observations. Privé *et al*. (2014) investigated sensitivity to unmanned aircraft observations for tropical cyclones using an OSSE approach and were able to show that different strategies could make a substantial difference to tropical cyclone forecasts. Peevey *et al*. (2018) followed this by exploring the impact of additional dropsonde observations using an OSSE approach on tropical cyclone forecasts. The number of cases studied was small, and difficult to infer a specific recommendation for SOFF, other than that the study suggested more observations would have been helpful. Kren, Cucurull and Wang (2020) explored sensitivity to flight tracks, again with an OSSE approach. Whilst again the number of cases, two, was too small to infer conclusions on the usefulness of changed tracks, it did show a sensitivity to changes, which suggested that with more extensive studies it may be possible to define optimal flight track strategies. Again though, for the purposes of this study, no substantial conclusions can be drawn for SOFF from that paper. The only study looking at upper air radiosonde data that we could find was Prive et al (2014) that looked at the sensitivity to providing additional launches at 06/18 Z, again using an OSSE framework, and found a small benefit. These past studies give only limited insight into the questions SOFF is asking. Therefore, the new experiments proposed will provide improved understanding of the impact of surface-based observations to help inform SOFF investments. We choose to use the EDA approach because it is much simpler and more affordable than a full OSSE but offers potentially more insight than local methods such as the 1D-Var approach.

1.4. Summary

All observing systems (surface-based observations and data from various satellite instruments) provide significant positive impact for at least some aspects of the numerical weather prediction (NWP) system. The results confirm the overall complementarity of the global observing system (Bormann et al, 2019).

Existing studies show that surface-based observations significantly impact forecast accuracy, especially in areas with sparse data.

- High impact of surface-based observations: Despite being fewer in number compared to satellite data, surface-based observations have a strong influence on forecast accuracy in many regions. It is also known that, in general, adding one new surface-based observation in a data sparse region has more impact than adding a similar new surface-based observation in a data-rich region.
- 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.
- Importance of both surface land and upper air stations: while both are critical elements of the global observing system, the relative importance of surface land

versus upper-air observations varies across studies. More studies suggest that upper air observations have more impact, and there is a tentative acceptance that upper air observations have more impact than surface land observations.

- Significant geographical observation gaps: There is evidence that data gaps in Africa (in particular regions such as East Africa, the Rift Valley, and the Horn of Africa), parts of the Pacific and Atlantic Oceans, the Arctic Ocean and Antarctica, need to be reduced for improved forecast accuracy.
- More scientific studies on the impact of surface-based observations are required. Most studies in the literature explore satellite data impact, studies on surface-based data are more limited. Existing studies on the impact of existing observations do not provide the level of granularity required to fully guide SOFF investment priority decisions.

In summary, there is substantial evidence that investing in additional surface-based observations is highly beneficial.

2. Proposed additional scientific studies: scenarios for SOFF-tailored Observation System Experiments

To assess the potential impact of additional surface-based observations, several scenarios are proposed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and WMO:

- Baseline: Current observing system.
- Ensemble of Data Assimilations Scenarios: Adding simulated surface land and upper-air observations for different country groupings (LDCs, SIDS, LMIC, all ODA eligible countries, FCS) and regions (Africa, Pacific).
- Observation System Experiments Scenarios: Simulating the absence of surface land and upper-air observations.

The baseline uses the current observing system. This means new observations must bring new information to have an impact. The scenarios then measure how much forecast error is reduced when we add simulated additional observations for a range of options.

With the baseline being the current observing system, the specific scenarios that will be studied are listed below:

a) Ensemble of Data Assimilations (EDA): Baseline plus simulated surface land and upper air observations for LDCs and SIDS (SOFF eligible for investment and compliance support as of today);

b) EDA: (a) plus simulated surface and upper air observations for Lower Middle-Income Countries (LMICs);

c) EDA: ODA countries and SIDS baseline plus simulated surface and upper air observations;

d) EDA: As (c) but for upper air observations only;

e) (a) plus marine simulated observations in same countries;

f) EDA: Baseline plus surface land observations for FCS states;

g) EDA: Baseline plus upper air observations for FCS states;

h) To be confirmed: EDA: Baseline plus simulated surface land and upper air observations in the Pacific;

i) To be confirmed: EDA: Baseline plus simulated surface land and upper air observations for Africa;

j) OSE: No real surface or upper air;

k) EDA: with and without surface land and upper air for existing observations.

As a means of calibration, an experiment will be run removing real surface-based observations to measure the current impact of surface-based observations. This can then be compared to an EDA study adding simulated equivalents. It will show how well the method replicates actual impact of real observations. This will provide confidence that results for simulated observations are realistic, as they replicate impact of real observations.

This report is exclusively on the value of observations to weather forecasts. There is a strong link between observation requirements for weather and climate. Therefore, whilst the study does not directly inform climate monitoring and prediction, the results may also give some insight into potential benefit also for climate.

References

Battaglia A., Rumi E., Reeves, R. et al. 2023: Surface Air-Pressure Measurements from Space Using Differential Absorption Radar on the right wing of the 60 GHz oxygen band. *ESS Open Archive* . November 08, 2023. DOI: 10.22541/essoar.169945274.46101502/v1

Ben-Bouallegue, Z., Clare, M.C., Magnusson, L., Gascon, E., Maier-Gerber, M., Janousek, M., Rodwell, M., Pinault, F., Dramsch, J.S., Lang, S.T. and Raoult, B., 2023. The rise of data-driven weather forecasting. *arXiv preprint arXiv:2307.10128*.

Bormann N, H. Lawrence and J. Farnan, (2019): Global observing system experiments in the ECMWF assimilation system, ECMWF Tech Memo 839. https://www.ecmwf.int/en/elibrary/80953-global-observing-system-experiments-ecmwf-assimilation-system. DOI 10.21957/sr184iyz.

Boukabara, S., and Coauthors, 2018: Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Perfect Observations Simulation Validation. *J. Atmos. Oceanic Technol.*, 35, 207–226, https://doi.org/10.1175/JTECH-D-17-0077.1.

Centurioni, L. R., Hornayi, A., Cardinali, C., Charpentier, E., and Lumpkin, R. (2017). A global observing system for measuring sea level atmospheric pressure: effects and impacts on numerical weather prediction. *Bull. Am. Meteorol. Soc* 98, 231–238. doi: 10.1175/BAMS-D-15-00080.1

Chambon, P., Mahfouf, J.F., Audouin, O., Birman, C., Fourrié, N., Loo, C., Martet, M., Moll, P., Payan, C., Pourret, V. and Raspaud, D., 2023. Global observing system experiments within the météo-france 4D-Var data assimilation system. *Monthly Weather Review*, *151*(1), pp.127-143.

Deblonde, G. and English, S., 2003. One-dimensional variational retrievals from SSMIS-simulated observations. *Journal of Applied Meteorology and Climatology*, *42*(10), pp.1406-1420.

Demortier, A., Mandement, M., Pourret, V. and Caumont, O., 2023. Assimilation of surface pressure observations from personal weather stations in AROME-France. *Natural Hazards and Earth System Sciences Discussions*, *2023*, pp.1-30.

Dumelow, R., 2004: Observing System Experiments using the Met Office global model, in Proceedings of the Third WMO workshop on the impact of various observing systems on Numerical Weather Prediction, Alpbach, Austria 9-12 March 2004, WMO/TD No. 1228, 95-111.

Falk U., Smart S. Wetterhall F., Quintino T., Peuch V.-H., 2023: Towards using unconventional observations. ECMWF newsletter 174. Winter 2023.

ECMWF (2021): Annual Seminar: Observations https://events.ecmwf.int/event/217/

English S.J., C. Poulsen, R. Renshaw (1999): The impact of assimilating ATOVS 1D-Var retrievals over Siberia. FRTR.269. Obtainable form the Met Office library (https://library.metoffice.gov.uk)

English S.J. (2018): Impact studies by ECMWF, including OSEs vs FSOI. CGMS-46, Agenda Item 5, WP13. https://www.cgms-info.org/agendas/agendas/CGMS-46.

Eyre, J.R., 1990. The information content of data from satellite sounding systems: A simulation study. *Quarterly Journal of the Royal Meteorological Society*, *116*(492), pp.401-434.

Eyre J.R. 2021: Observation impact metrics in NWP: A theoretical study. Part I: Optimal systems, Quarterly Journal of the Royal Meteorological Society, 10.1002/qj.4123, 147, 739, (3180-3200), (2021).

Eyre, J.R. 2023: Observation impact metrics in NWP: A theoretical study. Part II: Systems with suboptimal observation errors. Quarterly Journal of the Royal Meteorological Society, 1-9. Available from: https://doi.org/10.1002/qj.4614

Harnisch, F., Healy, S.B., Bauer, P. and English, S.J., 2013. Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations. *Monthly Weather Review*, *141*(12), pp.4395-4413.

Healy, S.B. 2013: Surface pressure information retrieved from GPS radio occultation measurements. Q.J.R. Meteorol. Soc., 139: 2108-2118. https://doi.org/10.1002/qj.2090

Ingleby N.B., M. Rodwell, L. Isaksen, 2015: Global radiosonde network under pressure. https://www.ecmwf.int/en/newsletter/149/meteorology/global-radiosonde-network-under-pressure

Ingleby, B., and L. Isaksen (2018): Drifting buoy pressures: Impact on NWP. Atmos. Sci. Lett., 19, https://doi.org/10.1002/asl.822.

Ingleby B., 2021: ECMWF blog March 2021. https://www.ecmwf.int/en/about/media-centre/science-blog/2021/importance-additional-surface-based-observations-data-sparse

Ingleby, B., B. Candy, J. Eyre, T. Haiden, C. Hill, L. Isaksen, D. Kleist, F. Smith, P. Steinle, S. Taylor, W. Tennant, and C. Tingwell, 2021: The impact of COVID-19 on weather forecasts: a balanced view. Geophys. Res. Let., https://doi.org/10.1029/2020GL090699.

Ito, K., and Coauthors, 2018: Analysis and forecast using dropsonde data from the inner-core region of tropical cyclone Lan (2017) obtained during the first aircraft missions of T-PARCII. *SOLA*, 14, 105–110, https://doi.org/10.2151/sola.2018-018.

James, E.P. and Benjamin, S.G., 2017. Observation system experiments with the hourly updating Rapid Refresh model using GSI hybrid ensemble–variational data assimilation. *Monthly Weather Review*, *145*(8), pp.2897-2918.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus*, 64A, 18462, https://doi.org/10.3402/tellusa.v64i0.18462.

Kitchen C., Bennett V. and Chantry M. (2023): The growing role of machine learning at ECMWF https://www.ecmwf.int/en/about/media-centre/news/2023/growing-role-machine-learning-ecmwf

Kren AC, Cucurull L, Wang H. Addressing the sensitivity of forecast impact to flight path design for targeted observations of extratropical winter storms: A demonstration in an OSSE framework. Meteorol Appl. 2020; 27:e1942. https://doi.org/10.1002/met.1942

Kull,D.W., L-P. Riishojgaard, J. Eyre, R.A. Varley (2021): The Value of Surface-based Meteorological Observation Data (English). Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/192461614151036836/The-Value-of-Surface-based-Meteorological-Observation-Data.

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

Laroche, S. and Poan, E.D., 2022. Impact of the Arctic observing systems on the ECCC global weather forecasts. *Quarterly Journal of the Royal Meteorological Society*, *148*(742), pp.252-271.

Lawrence H, Bormann N, Sandu I, Day J, Farnan J, Bauer P. (2019): Use and impact of Arctic observations in the ECMWF Numerical Weather Prediction system. *Q J R Meteorol Soc*; 145: 3432–3454. https://doi.org/10.1002/qj.3628

Liu, Z., Snyder, C., Guerrette, J.J., Jung, B.J., Ban, J., Vahl, S., Wu, Y., Trémolet, Y., Auligné, T., Ménétrier, B. and Shlyaeva, A., 2022. Data assimilation for the Model for Prediction Across Scales–Atmosphere with the Joint Effort for Data assimilation Integration (JEDI-MPAS 1.0. 0): EnVar implementation and evaluation. *Geoscientific Model Development*, *15*(20), pp.7859-7878.

Magnusson, L., 2017: Diagnostic methods for understanding the origin of forecast errors. Q.J.R. Meteorol. Soc, 143: 2129-2142. https://doi.org/10.1002/qj.3072.

Pauley, P.M., Ingleby, B., 2022: Assimilation of In-Situ Observations. In: Park, S.K., Xu, L. (eds) Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV). Springer, Cham. https://doi.org/10.1007/978-3-030-77722-7_12

Peevey, T. R., J. M. English, L. Cucurull, H. Wang, and A. C. Kren, 2018: Improving Winter Storm Forecasts with Observing System Simulation Experiments (OSSEs). Part I: An Idealized Case Study of Three U.S. Storms. Mon. Wea. Rev., 146, 1341–1366, https://doi.org/10.1175/MWR-D-17-0160.1.

Prates C., T. Pröscholdt, L. Nunes, A. Abouelenein, T. Kleinert, T. Oakley 2023: Monitoring the Global Basic Observation Network. https://www.ecmwf.int/en/newsletter/177/news/monitoring-global-basic-observation-network

Privé, N.C., McGrath-Spangler, E.L., Carvalho, D., Karpowicz, B.M. and Moradi, I., 2023. Robustness of Observing System Simulation Experiments. *Tellus A: Dynamic Meteorology and Oceanography*, 75(1), p.309–333.DOI: https://doi.org/10.16993/tellusa.3254

Privé, N.C., M. McLinden, B. Lin, I. Moradi, M. Sienkiewicz, G. M. Heymsfield, and W. McCarty, 2023: Impacts of Marine Surface Pressure Observations from a Spaceborne Differential Absorption Radar Investigated with an Observing System Simulation Experiment. *J. Atmos. Oceanic Technol.*, 40, 897–918, https://doi.org/10.1175/JTECH-D-22-0088.1.

Privé, N. C., Y. Xie, S. Koch, et al. R. Atlas, S. J. Majumdar, and R. N. Hoffman. 2014. An Observing System Simulation Experiment for the Unmanned Aircraft System Data Impact on Tropical Cyclone Track Forecasts Monthly Weather Review 142 (11): 4357-4363 [10.1175/MWR-D-14-00197.1]

Prive, N. C., R. M. Errico, and K.-S. Tai. 2014. The Impact of Increased Frequency of Rawinsonde Observations on Forecast Skill Investigated with an Observing System Simulation Experiment Mon. Weather Rev. 142 (5): 1823-1834 [10.1175/MWR-D-13-00237.1]

Randriamampianina, R., Bormann, N., Køltzow, M.A., Lawrence, H., Sandu, I. and Wang, Z.Q., 2021. Relative impact of observations on a regional Arctic numerical weather prediction system. *Quarterly Journal of the Royal Meteorological Society*, *147*(737), pp.2212-2232.

Schäfler, A., and Coauthors, 2018: The North Atlantic waveguide and downstream impact experiment. *Bull. Amer. Meteor. Soc.*, 99, 1607–1637, https://doi.org/10.1175/BAMS-D-17-0003.1.

Valmassoi, A., Keller, J.D., Kleist, D.T., English, S., Ahrens, B., Ďurán, I.B., Bauernschubert, E., Bosilovich, M.G., Fujiwara, M., Hersbach, H. and Lei, L., 2023. Current challenges and future directions in data assimilation and reanalysis. *Bulletin of the American Meteorological Society*, *104*(4), pp.E756-E767.

Weissmann, M., and Coauthors, 2011: The Influence of Assimilating Dropsonde Data on Typhoon Track and Midlatitude Forecasts. *Mon. Wea. Rev.*, 139, 908–920, https://doi.org/10.1175/2010MWR3377.1.

WMO (2020): Seventh WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction. Edited by Sid Boukabara (NOAA), SeiYoung Park (KMA),

WMO (2024): 8th WMO Workshop on the Impact of Various Observing Systems on Numerical Weather Prediction and Earth System Prediction. SMHI, Norrköping, Sweden, 27 to 30 May 2024. Available from https://community.wmo.int/en/meetings/8th-wmo-impact-workshop-home.

Yamazaki, A., T. Miyoshi, J. Inouem, T. Enomoto and N. Komori (2021). EFSO at Different Geographical Locations Verified with Observing System Experiments, *Weather and Forecasting*, *36*(4), 1219-1236. Retrieved Sep 16, 2021, from https://journals.ametsoc.org/view/journals/wefo/36/4/WAF-D-20-0152.1.xml.