

Use of Observing System Simulation Experiments (OSSEs) at NOAA

9 April 2019

Prepared by the OSSE Task Force

1. Introduction

The NOAA Scientific Advisory Board (SAB) identified the review of the use of Observing System Simulation Experiments (OSSEs) as one of its key activities, and tasked its Environmental Information and Service Working Group (EISWG) to lead this effort in collaboration with the Climate Working Group (CWG). EISWG submitted the OSSE work plan to SAB in August 2018 and finalized the team membership in December 2018, including

- EISWG Members: Bill Hooke, Ron Birk, Bob Weller, Xubin Zeng (Chair)
- SAB members: Eugenia Kalnay and Susan Avery (SME)
- CWG members: Joellen Russell, Fuqing Zhang, Raghu Murtugudde
- NOAA liaison: Lidia Cucurull and Bob Atlas, NOAA AOML
- Domain experts: Fred Carr (note that other people mentioned here, such as Bob Atlas and Eugenia Kalnay, are also domain experts)
- External Agency Partners: Derek Posselt (NASA JPL)

While we don't have a member from Naval Research Laboratory (NRL), Dan Tyndall and his colleagues (N. Baker, D. Flagg, C. Barron, M. Carrier, S. Smith, D. Allen, and K. Hoppel) helped us by providing the summary of OSSEs at NRL (see the Supplementary Material).

The objectives of this report are:

- to review the use of OSSEs in NOAA, Navy, NASA, and elsewhere
- to develop options for NOAA to consider current and future research and development (R&D) work in this area, such as the combination of OSSEs with EFSO (which is being carried out at AOML with University of Maryland collaboration, and is expected to strongly enhance and accelerate the current abilities of OSSEs).

The deliverable is:

- a short white paper that will review the use of OSSEs in NOAA, Navy, NASA, and elsewhere; and develop recommendations for NOAA to consider and provide rationales for each recommendation made.

Individual write-ups were provided by team members from late December 2018 to early February 2019 (see Supplementary Material), and a face-to-face meeting (with X. Zeng, J. Snow, B. Hooke, R. Birk, J. Russell, F. Carr, and D. Posselt present) was held in January 2019. Initial thoughts from the preliminary report were presented to the SAB telecon in late February 2019, and the draft white paper was presented to the EISWG in early April 2019. The White Paper was finalized through several iterations among the OSSE Task Force members and with input from EISWG members, and it was submitted to SAB in late April 2019.

2. Findings on the use of OSSEs at NOAA and elsewhere

An OSSE is a modeling experiment used to evaluate the impact of new observing systems on operational forecasts when actual observational data are not available. OSSEs are done:

- to find out if a new observing system will add value to NWP analyses and forecasts;
- to make design decisions for a new observing system; and

- to investigate the behavior of data assimilation systems in an environment where the system's behavior is known.

Since 1980s, the current methodology used for rigorous OSSEs has been accepted nationally and internationally as the way in which OSSEs should be conducted in order to provide credible results. Since that time, extensive OSSEs have been conducted, first at NASA/GSFC, and later at NOAA/AOML in collaboration with operational data assimilation centers, private enterprise, and academic partners. These OSSEs determined correctly the quantitative potential for several proposed satellite observing systems to improve weather analysis and prediction prior to their launch, evaluated trade-offs in orbit configurations, coverage and accuracy for space-based observing systems, and were used in the development of the methodology that led to the first beneficial impacts of satellite surface winds on numerical weather prediction. Today, OSSEs and related capabilities exist at NOAA, NASA, NRL, universities, the private sector, and the Federally Funded Research and Development Centers (FFRDCs) (see the partial list in the Supplementary Material).

Since 2014, OSSEs in NOAA have been performed under NOAA's Quantitative Observing System Assessment Program (QOSAP). QOSAP coordinates the assessment of the impact of current and new observations across the different NOAA Line Offices and it uses observing system experiments (OSEs, to test value of existing, rather than future, observing systems), forecast sensitivity observation impact (FSOI) and ensemble FSOI (EFSOI), and OSSEs as effective techniques to evaluate the impact of the different observation types. QOSAP's primary objective is to improve quantitative and objective assessment capabilities to evaluate operational and future observation system impacts and trade-offs to assess and to prioritize NOAA's observing system architecture. More specifically, QOSAP's main focuses are (1) to increase NOAA's capacity to conduct quantitative observing system assessments, (2) to develop and use appropriate quantitative assessment methodologies, and (3) to inform major decisions on the design and implementation of optimal composite observing systems.

Under QOSAP, a state-of-the-art global OSSE system, an advanced Hurricane OSSE System, and an internationally recognized first of its kind rigorous Ocean OSSE System were developed. For global NWP, a state-of-the-art global OSSE system based on the NASA Cubed Sphere (FV3) at 7 km resolution nature run (with the model output assumed to closely represent the true environmental conditions) was developed to allow observation impact assessments at higher horizontal resolution. QOSAP also began acquisition and initial testing of a new 9-km horizontal resolution global nature run provided by ECMWF. Development of regional OSSE systems for high impact weather and air quality were initiated, and a 2-km state-of-the-art basin scale nature run has been developed.

Using these systems, a significant number of OSEs and OSSEs in both global and regional (tropical cyclone) systems for multiple existing and proposed observing systems were performed and many of these have since been published in the refereed literature. Additionally, QOSAP conducted OSSEs related to the role of ocean observations in hurricane prediction. In particular, QOSAP met the deadlines to complete OSSEs with GNSS-RO and Geo-HSS required by U.S. Congress under the Weather Law H.R. 353. Finally, QOSAP began the process to develop the quantitative assessments capability to meet the needs of NOS and NMFS, and OSSE capabilities for other ocean basins, coastal oceans, and for climate are under development. These capabilities are summarized in Figure 1.

Besides NOAA, NASA has been conducting OSSEs for decades, primarily by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center (GSFC). The goal is to determine how much additional information is provided by a new set of measurements, relative to the current global observing system. This is consistent with NASA's aim of providing accurate and complete characterization of the state of Earth's atmosphere, oceans, land surface, and cryosphere. The primary product produced by NASA's modeling and data assimilation infrastructure is the Modern-Era Retrospective-analysis for Research and Applications (MERRA). GMAO also conducts research into how to properly calibrate an OSSE and has also produced a global mesoscale-resolving Nature Run (also used by NOAA). The GMAO OSSE system consists of the NASA Global Earth Observing System (GEOS)

model and the Gridpoint Statistical Interpolation (GSI) data assimilation system. Note that the GSI is also used in the NOAA NCEP operational forecast system.

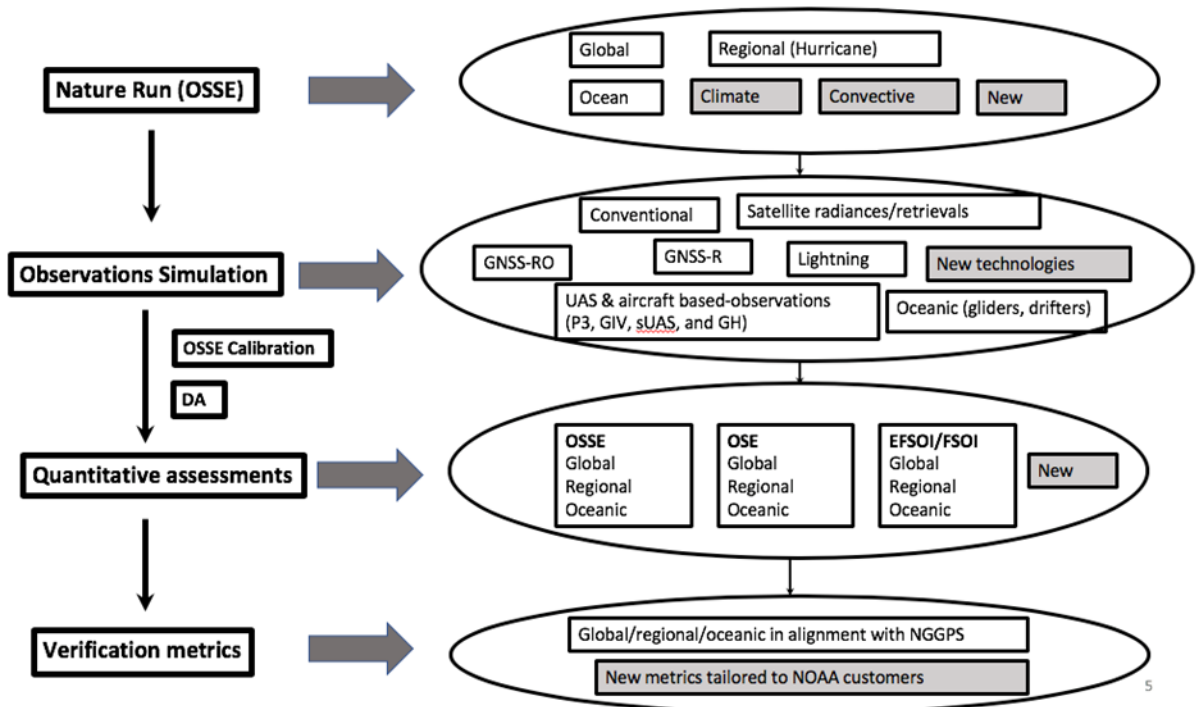


Figure 1. Diagram showing the different components of the OSSE system at NOAA AOML.

In general, Earth observations from NASA have two purposes: (1) accurate characterization of the Earth’s atmosphere, oceans, cryosphere, and land surface, and (2) scientific discovery of the processes that drive the evolution of the Earth system, and the linkages among the components of the system. Besides traditional forecast OSSEs, NASA has also done sampling OSSE (to address the question of whether a set of measurements is able to see a feature of interest) and retrieval OSSE (to quantify the degree to which prospective measurements provide information on a geophysical quantity of interest).

The U.S. Navy requires meteorological and oceanographic information to characterize the environment to support global, regional, and tactical scale operations on time scales ranging from minutes to weeks. Because the battlespace environments the Navy operates in are often data sparse, investments in new observation types addressing insufficiently-sampled properties are critical. Recently, the Navy has requested estimates of impacts that potential observing systems would have on NWP forecasts and tactical decision aids before fully investing in the systems.

These estimates can be computed using an OSSE; however, the traditional OSSE methodology can be costly in both personnel and computational resources associated with the production of a nature run as well as the simulation of both new observations and existing observations from the global observing system. Instead of running traditional OSSEs to estimate observation impacts, the Naval Research Laboratory (NRL) has run several variants of the methodology in recent years to derive similar statistics (Fig. 2) for

- the Coupled Ocean-Atmosphere Mesoscale Prediction System using the NCEP Global Forecast System (GFS) analysis fields (to replace the Nature run),
- the Navy Coastal Ocean Model (NCOM) (by simulating observations from a Nature run or using NCOM model data from a different year - but the same month and day), and
- the Navy Global Environmental Model to study impacts of potential observations (e.g., stratospheric ozone) on middle atmosphere prediction.

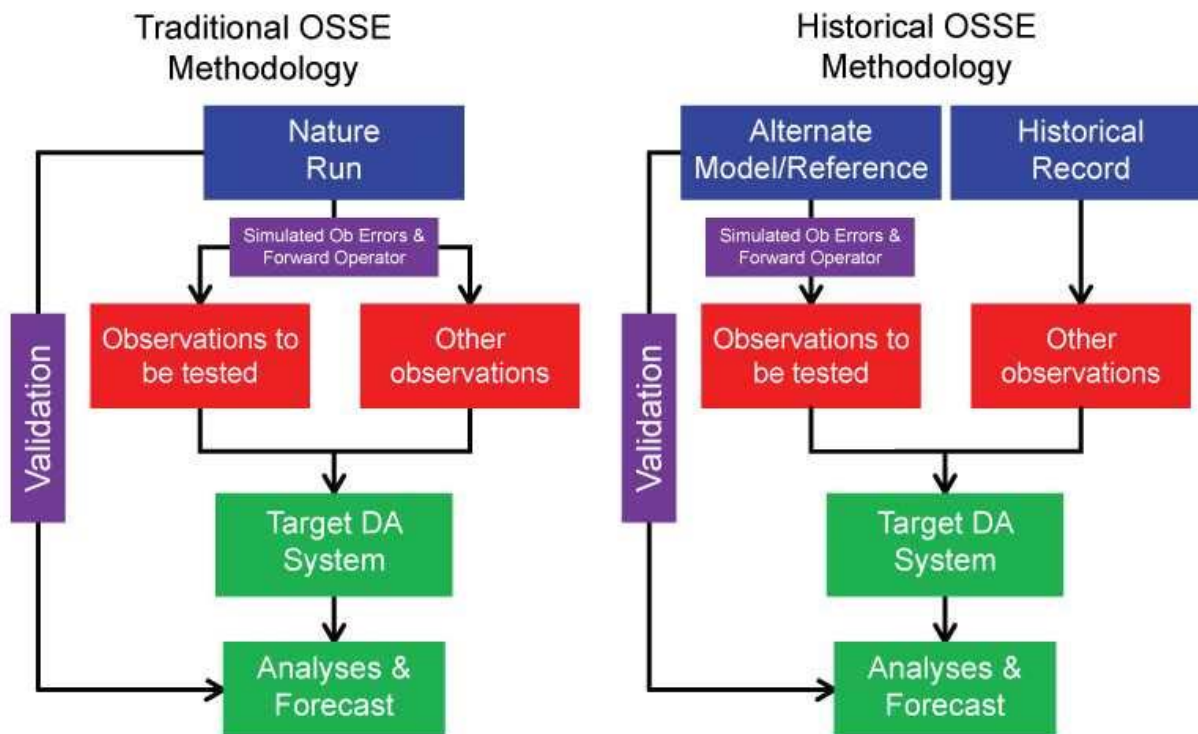


Figure 2. Schematic depiction of traditional OSSE methodology (left) and the historical OSSE methodology (right)

The private sector recognizes the value of developing, evolving, and applying OSSEs to inform decisions on investments in observing system capabilities. Assessments are conducted to inform plans and designs for commercial sector observing systems, including making the case to investors for the value of the remote sensing systems. Assessments are also used to inform decisions on design alternatives for government systems, by both private sector developers and FFRDCs assessing the value of alternatives. A representative list of private sector organizations engaged in OSSEs is provided in the supplemental material.

Besides OSSEs, operational or research data assimilation systems (e.g., at NCEP) can provide real-time assessments of the sensitivity of the final analysis to the individual observations used in the analysis (although they are usually grouped by observation system). This is known as the Forecast Sensitivity to Observations (FSO) method. This was first done with adjoint data assimilation systems – because of the use of the tangent linear model, this approach is limited to short-range forecast (1-2 days) impacts. FSOs can also be done using Ensemble Kalman Filter data assimilation systems, known as EFSOs (e.g., at the University of Maryland). EFSO uses ensemble perturbations to evaluate during the 6hr forecasts whether each observation is beneficial or detrimental. Proactive Quality Control (PQC) then deletes, for example, the 10% most detrimental observations, resulting in large forecast improvements, and a collection of detrimental observations that can facilitate improving the observation algorithms. Therefore, combining OSSEs with EFSO will provide much more information about each observing system, make OSSE+EFSO/PQC much more effective and useful than OSSE alone.

3. Discussions on the values and limitations of OSSEs

As discussed in Section 2, the values of OSSEs have been demonstrated at NOAA, NASA, and NRL in evaluating the impact of new observing systems on operational forecasts when actual observational data are not available. OSSEs are also valuable for testing new data assimilation methodologies and for observation targeting strategies. They can be further enhanced by

- combining OSSEs with EFSO (or FSO) – as OSSEs can see the forecast impact of one particular (proposed) observing system, but it won't see the individual impact of all the other (simulated) observing systems, which EFSOs and FSOs can do,
- using different approaches to replace the Nature run – as Nature run may have specific deficiencies,
- using variety of OSSEs (for forecasting, sampling, and retrievals) – as forecast OSSEs may not be able to address the impact of observations in answering specific science questions, and
- using OSEs for current observation systems.

While most OSSEs are done for global satellite systems, they can be used to assess new observing systems for regional scales as well. For instance, on the storm-scale, OSSEs have been done at University of Oklahoma to assess the value of using dual-polarization radar data in NWP, as well as to assess different scanning strategies and network configurations. On the continental U.S. scale, OSSEs can be used to assess the value of increasing the density of vertical profiling systems, as recommended by the National Academies "Network of Networks" Report. Besides OSSEs/OSEs, several National Academies reports have recommended that new observing systems be deployed in regional testbeds for evaluation (including urban testbeds) for evaluation, before investing in a nation-wide system.

However, the reliability and effectiveness of OSSEs depend critically on the data assimilation methodology and the forecast models. In particular, the relative impacts of different observation systems may depend critically on the data assimilation system. For example, the current NCEP operational data assimilation system has demonstrated no quantifiable impacts of any all-sky (cloudy/rainy) satellite radiances in the operational model performance (based on the recent NCEP presentations) while the more advanced ECMWF data assimilation and prediction system now puts all-sky radiance (minus clear-sky) as the most impactful sources of observations (based on the ECMWF presentation at the AMS Annual Meeting in January 2019). Furthermore, OSSEs should state the time period for which the results are valid, depending on the use of current or future observing systems.

The use of extreme events (e.g., a major hurricane event) or their nature runs as truth for OSSEs should be avoided, as the skill scores for any operational NWP models are judged by a large number of cases or seasons, and by many metrics, not a single event. In general, individual events in real world or in nature runs can have case-dependent and flow dependent predictability, which will have significant impacts on using certain observing systems. A continuous long "nature" run, on the other hand, is likely to be drifting away from the true nature, given the unavoidable yet still significant errors in the model physics or in the boundary conditions or forcings.

While OSSEs for the global atmosphere are relatively mature, further development for ocean OSSEs is needed. Despite its major role in the earth system, the ocean is sparsely observed. The lack of observations stems in part from the technical challenges of sustaining observations of the ocean and also from the cost of maintaining observing arrays and networks across the ocean basins. In this context, OSSE, especially combined with EFSO, are a valuable tool that can be used to inform those that fund ocean observations about the impact that specific observing elements have on model fields. Additional OSE efforts are also valuable. For instance, the large, cooperative European Union project AtlantOS will focus on a forward design for basin-scale in situ observations, with a quantitative focus informed by OSE/OSSE work.

However, a basic challenge for ocean OSEs and OSSEs is that because the ocean is sparsely sampled and since the ocean models do not capture all the modes and variability present, the realism of the models and conclusions about the impacts of observing system elements need to be questioned and considered with care. For instance, recent community efforts indicate that tropical Pacific OSE/OSSE studies are

expensive (usually) and often inconclusive, in large part due to the large systematic errors in models and dependence on parameterization assumptions. Therefore, multiple lines of evidence are encouraged to support detected sensitivity.

As mentioned earlier, OSSEs have great power for inexpensively and rapidly exploring the impact of the relative contributions made to NWP by a wide range of observing technologies -- and indeed providing insights into a number of observing configurations that might be prohibitively expensive and time consuming to develop by any other means. On the other hand, attention shouldn't be confined to OSSEs to the exclusion of other R&D, such as

- actual deployment and use of observing technologies in pilot programs and demonstration projects;
- complementing advances in NWP per se with corresponding improvements in mass risk communication and the use of new technologies such as data analytics and artificial intelligence;
- basic social science research toward similar ends;
- R&D in valuing weather information; and
- other avenues.

The opportunities – and the public stakes (with respect to health and safety and building resilience to hazards; development of renewable natural resources; and protecting the environment and ecosystems) – are so high and so urgent as to demand a national pursuit of all these diverse R&D and technology transfer paths in parallel, rather than in sequence or in isolation. More attention to OSSEs and development of their potential is needed, but in a manner balanced by additional attention to other opportunities across the board.

4. Recommendations on potential NOAA actions related to OSSEs

NOAA is mandated by the Weather Research and Forecasting Innovation Act of 2017 Section 107 to perform OSSEs. Indeed OSSEs have been successfully used in major decision-making in the past. For instance, there was a proposed data buy where NASA and NOAA were each required to spend \$150M to buy a particular type of data. A joint NOAA/NASA OSSE was performed to determine the data requirements for this observing system. It was determined that the minimum requirements to ensure a beneficial impact on weather prediction for this observing system could not be met, and the nation did not have to make this unnecessary expenditure.

Based on the findings in Section 2 and discussions in Section 3, here are our recommendations on potential NOAA actions related to OSSEs.

Recommendation 1: OSSE, OSE, FSO, EFSO research efforts should be coordinated nationally (e.g., sharing of software tools) to avoid duplication of effort (e.g., via the QOSAP program). These methods each have their pros and cons, and should all be used to assess the relative benefit of different observing systems. Besides full-scale OSSE experiments, simple experiments could also be very powerful (e.g., for sampling strategies and data value evaluation).

Recommendation 2: The OSSE development for earth system models (e.g., for sea ice prediction) needs to be accelerated. Furthermore, global 5 km (and preferably 3 km) Nature Run based on earth system models should be developed as the basis for a variety of OSSEs. This may require the purchase of new high-performance computers or the partnership with other agencies.

Recommendation 3: NCEP data assimilation and prediction system will continue to improve. OSSEs are used to evaluate the observational network likely decades ahead. Therefore, the choice of observations and investment decisions based on OSSEs need to explicitly consider the potential impact of deficiencies in the current data assimilation and prediction system.

Recommendation 4: Besides existing OSSE activities at NOAA, OSSEs should also be used to:

- assess the value of NOAA partnership in satellite remote sensing with foreign agencies (e.g., India) and the private sector (e.g., purchasing data from privately-launched satellites),
- assist the exploration of strategies for the most effective and efficient way to do sea ice prediction (observations, models, data assimilation). Should NOAA request ice-breakers? How many?
- compare the value of (polar, geostationary, small/cube) satellite network strategy (e.g., small number of large satellites versus large number of small and cube satellites) for weather and climate prediction, and
- do a gap analysis in NOAA; i.e., what are the greatest new observational needs? What combination of old and new systems will work best?

Recommendation 5: OSSEs have been primarily used to evaluate the impacts of observing systems and/or observation denial on forecast performance per se, that is, on the physical parameters, and treating all forecast locations, times, and circumstances as equal. But this idea should be extended to societal impacts, whether monetizable, or in terms of lives at stake, etc. In other words, there are national priorities (e.g., saving human race) where money does not matter, and there are priorities depending upon the constraint of financial resources. This could be a possible additional avenue of research. In an Earth system model where social systems and the built environment are included, one can imagine collecting human data or propagating just the physical earth system information through the social systems as well.

Indeed, while OSSEs provide quantitative analyses of future observing system impacts for a specific model, the effects on products that rely on that model can only be estimated qualitatively. The NOAA/NESDIS Technology, Planning and Integration for Observation (TPIO) division has developed a qualitative tool for assessing supporting investment decisions, called the NOAA Observing System Integrated Analysis (NOSIA-II), also known as NOAA's Value Tree. This Value Tree is based on the survey of subject matter experts across all NOAA Line Offices to gauge the impacts of Earth observation investments on NOAA's key products and services. Therefore the aforementioned OSSE, OSE, FSO, EFSO and PQC tools should be used in concert with the current NOSIA-II system to determine NOAA's future observing needs.

Finally, it should be emphasized that perhaps the greatest benefit of R&D on OSSEs is not so much the guidance they can provide by themselves with respect to any particular observing system development and deployment decision. Instead it's about the enriched perspective they provide about strategic approaches to investment in Earth observations, science, and services in support of the national agenda. There is an analogy to the famous Eisenhower quote "individual plans are worthless, but planning is vital."