Supplementary Material for the OSSE White Paper
Written contributions by individual Task Force members and colleagues at NRL

1. OSSE capabilities in the private sector (R. Birk, 12/31/2018, revised on 2/25/2019)

Universities
- University of Wisconsin-Madison – [Infrared Brightness Temperatures](#)
- University of Wisconsin-Madison – [Surface-based Boundary Layer Profiler](#)
- University of Miami – [Fraternal Twins Ocean](#)
- University of Miami – [Future OSSEs](#)
- Duke University - [Meridional Volume, Heat, and Freshwater Transports](#)
- Florida State University – [Laser Atmospheric Wind Sounder](#)
- University of Oklahoma – [Ensemble Square Root Kalman Filter](#)
- MIT - [Improving Numerical Weather Prediction](#)
- MIT – [Advanced Fisheries Management Information System](#)
- University of Maryland – [O2R/R2O Infrastructure](#)
- Pennsylvania State University – [HWRF Hurricane Ensemble](#)
- Pennsylvania State University – [Multiscale Tropical Weather Systems](#)
- Morgan State University – [Performance and Evaluation of GMAO OSSE](#)

Companies
- AER - [Water Vapor Profiles and Rainfall](#)
- AER – [Lidar Wind and Water Vapor Observations](#)
- AER – [Ocean Surface Winds from CYGNSS](#)
- AER and GeoOptics – [QuickOSSE for GNSS RO](#)
- Astra Space – [Ionospheric and thermospheric measurements](#)
- GeoOptics and AER – [Extreme Weather Study](#)
- Ball Aerospace -[Doppler Wind Lidar](#)
- Ball Aerospace – [Autocovariance Wind Lidar](#)
- Riverside – [Community Global OSSE Package (CGOP)](#)

FFRDCs
- NCAR/UCAR – [Atmospheric Composition and Remote Sensing Prediction](#)
- NCAR/UCAR - [ECMWF Nature Run](#)
- Aerospace – [GPSRO on HWRF Accuracy](#)
- Aerospace – [Forecast Sensitivity to Observations](#)

2. OSSEs for ocean observing and ocean science (R. Weller, 1/3/2019)

The ocean plays a major role in the earth system. It stores freshwater, heat, and carbon dioxide; and anomalous ocean conditions, such as the warm sea surface temperatures (SSTs) of El Niño events, drive anomalous changes in rainfall and temperature over the land. At the same time, the ocean's fisheries and mineral deposits are resources with large societal impact. In spite of the importance of the ocean, our ability to better understand and model the ocean, its ecosystems, and its roles in weather and climate are limited by the fact that 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, Observing System Simulation Experiments (OSSEs) 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 Observing System Evaluation (OSE) efforts are also valuable. The OSE's use runs of data assimilative models in which different elements of the actual observing system now in place are withheld to examine the sensitivity of the resultant model fields on these observing elements. The OSSE's use a model run, called the nature run, to yield a simulation of the ocean taken as reality, then sample the fields of the nature run to produce simulated observations. The impact of the simulated observations on a second data assimilative model run, the forecast run, is then investigated. 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.

In recent years, the ocean community has been engaged in examining the Tropical Pacific Observing System (TPOS) used to monitor the equatorial Pacific and provide data to models to forecast El Niño events. Funding and operational challenges had resulted in reduced observing capabilities, and concern over maintaining capabilities in the TPOS to support prediction and improved understanding gave rise to an international study of how best to evolve and maintain the TPOS. An early report of the TPOS 2020 effort (https://www.godae-oceanview.org//files/download.php?m=documents&f=141030123052-OSEvalTTabsSmith1BOral.pdf) noted:

"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 parameterisation assumptions. TPOS sponsors are placing high priority on OSE guidance and TPOS 2020 would be interested in exploring improved frameworks for conducting such studies. These frameworks would encourage multiple lines of evidence to support detected sensitivity. "

Fujii et al (2015) present results from OSE's of the TPOS and are a good example of an ocean OSE:

"To assess the relative roles of the TAO/TRITON and Argo data in constraining the upper ocean thermal structure and improving ENSO forecasts, four OSE runs were performed. The four runs and their designations include:

- CTL: no ocean profiles assimilated,
- ALL: all ocean profiles assimilated,
- noMoor: all ocean profiles assimilated except the mooring profiles,
- noArgo: all ocean profiles assimilated except Argo.

The model SSH from each OSE run is validated against independent satellite altimeter SSH observations."

Interest in using OSEs and OSSEs to guide ocean observing is also seen in a number of other areas. The large, cooperative European Union project AtlantOS. "One of the aims of AtlantOS is to achieve a transition from a loosely-coordinated and fragmented set of existing ocean observing activities, into a system that is sustained and sustainable, efficient, and fit-for-purpose." (https://www.atlantos-h2020.eu/ga3/mid-term-project-review-of-atlantos-in-the-north-and-south-atlantic-activities-with-focus-on-ecosystem-and-climate-wp-5-meeting/). A workshop planned for February 2019 "will focus on a forward design for AtlantOS basin-scale in situ observations, with a quantitative focus informed by OSE/OSSE work."

(http://goosocean.org/index.php?option=com_oe&task=viewEventRecord&eventID=2397). Halliwell et al.(2017) explore whether or not the collection of additional ocean observations, in this case expendable ocean profilers of temperature and salinity and drifting thermistor strings to capture upper ocean temperature structure, lead to improvements in the initialization of the ocean model used in Tropical Cyclone forecasts. Their approach uses the HYbrid Coordinate Ocean Model (HYCOM) for both the nature run and the forecast model run, though at different resolutions ("0.08° horizontal Mercator mesh and 26 vertical layers versus 0.04° horizontal Mercator mesh and 35 vertical layers"). The different resolutions were chosen to both attempt to get more realistic representation of the ocean mesoscale variability in the nature run and to introduce some differences between the model runs. Halliwell et al. (2017) show that the additional observations reduce biases in the upper ocean temperature fields.

Thus, OSSE's are in use in the ocean community and seen as a valuable tool to guide further development of ocean observing systems. However, the limited realism of existing ocean models and their need to parameterize unresolved process present challenges and in the data sparse ocean, care is required in using OSSEs.

References:

Fujii, Y., Cummings, J. , Xue, Y. , Schiller, A. , Lee, T. , Balmaseda, M. A., Rémy, E. , Masuda, S. , Brassington, G. , Alves, O. , Cornuelle, B. , Martin, M. , Oke, P. , Smith, G. and Yang, X. (2015), Evaluation of the Tropical Pacific Observing System from the ocean data assimilation perspective. Q.J.R. Meteorol. Soc., 141: 2481-2496. doi:10.1002/qj.2579
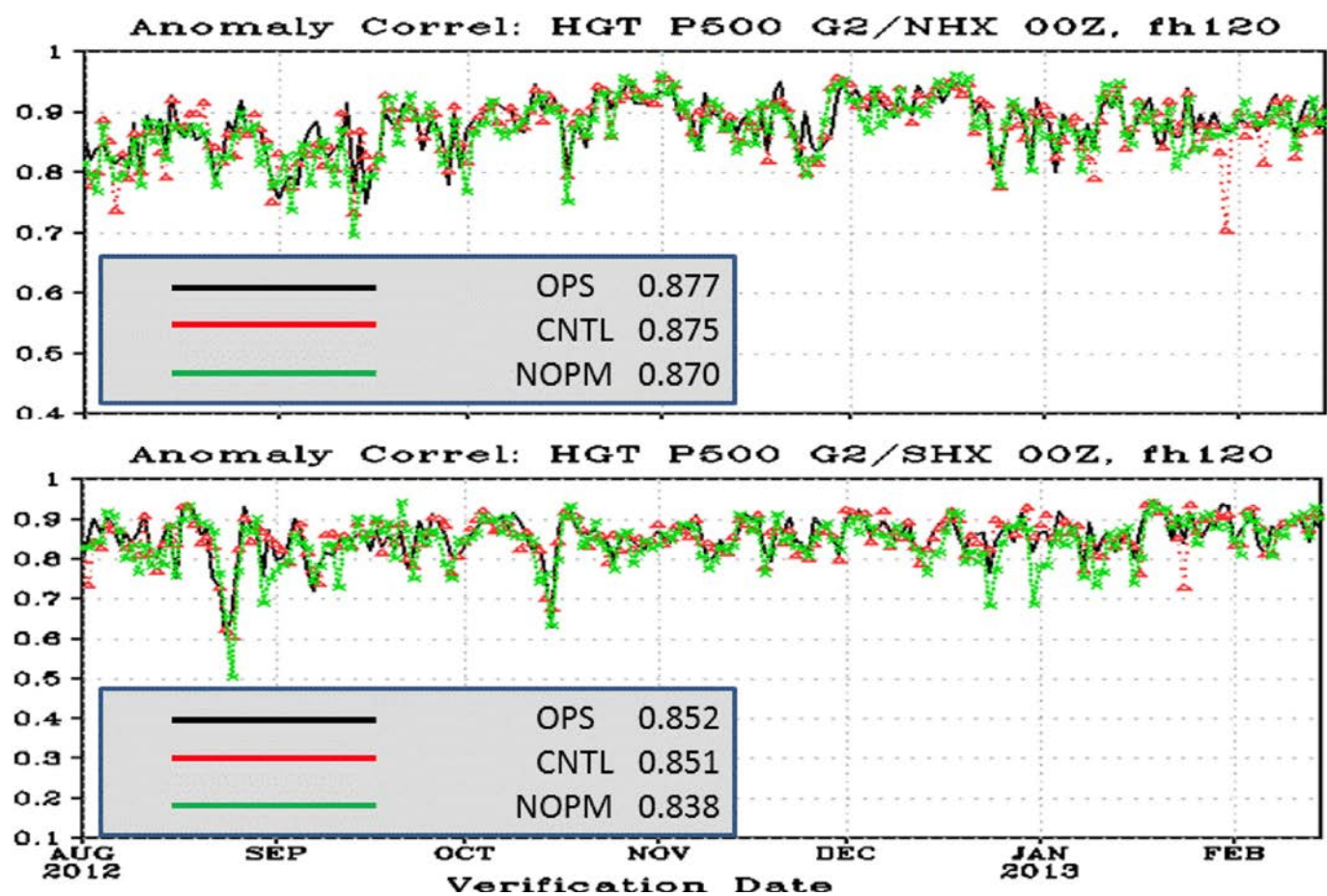
Halliwell, G. R., M. Mehari, L. K. Shay, V. H. Kourafalou, H. Kang, H.-S. Kim, J. Dong, and R. Atlas (2017), OSSE quantitative assessment of rapid-response prestorm ocean surveys to improve coupled tropical cyclone prediction, *J. Geophys. Res. Oceans*, 122, 5729–5748, doi:10.1002/2017JC012760.

3. OSSE + EFSO (E. Kalnay, 1/3/2019) – prepared in three slides

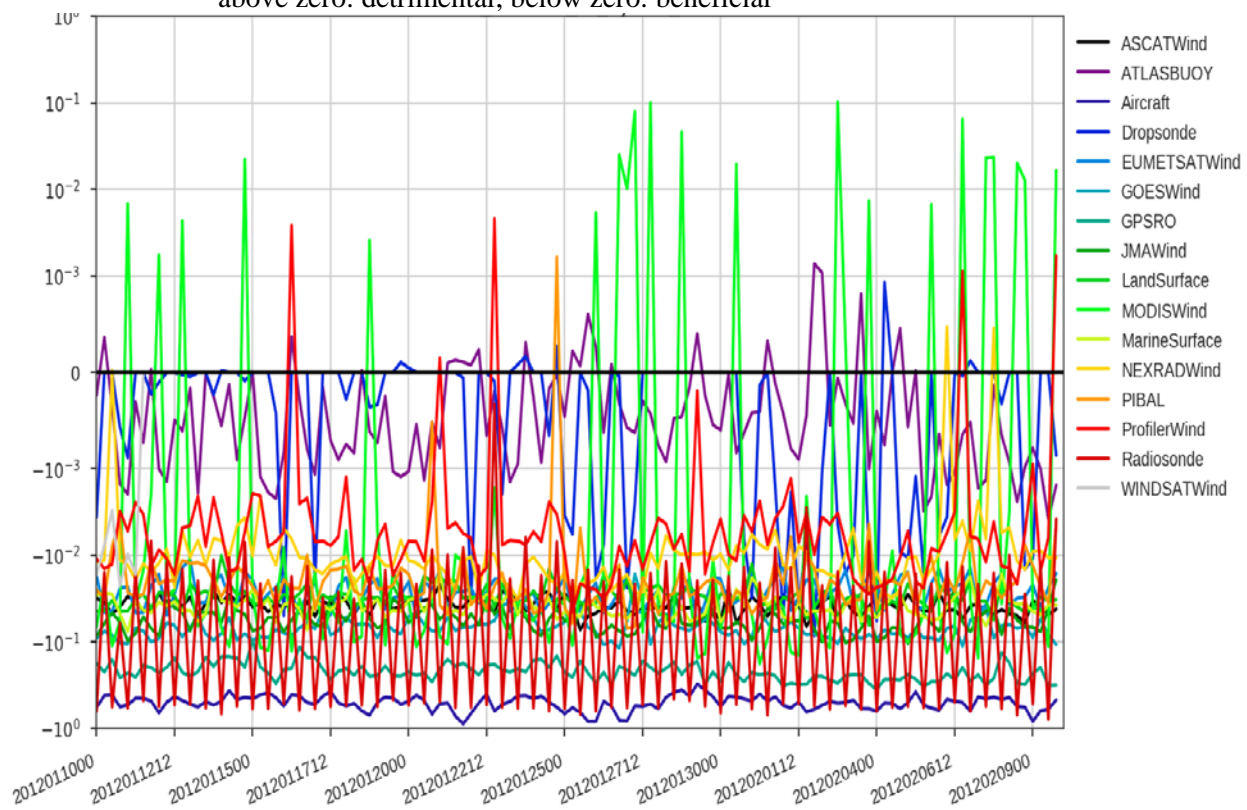OSSEs + EFSO are much more powerful than OSSEs alone!

- OSSEs have a lot of potential, but cannot separate the impact of the different observing systems on the forecasts.
- Ensemble Forecast Sensitivity to Observations (EFSO) can evaluate during the 6hr forecasts whether each observation is beneficial or detrimental.
- Combining OSSEs with EFSO (Chen and Kalnay, 2018) will provide much more information about each observing system, as shown in the next two slides.
- This will make OSSE+EFSO much more effective and useful!

In OSSEs or OSEs the forecast impact of all the instruments are mixed together. OSSE+EFSO: Monitors the impact of every instrument!

Anomaly Correl: HGT P500 G2/NHX 00Z, fh120

OPS	0.877	
CNTL	0.875	
NOPM	0.870	

Anomaly Correl: HGT P500 G2/SHX 00Z, fh120

OPS	0.852	
CNTL	0.851	
NOPM	0.838	

Verification Date

Lord et al., BAMS, 2016: OSE simulating failure of JPSS (all PM satellites)

06hr System Total Error Impact for each instrument (J/kg).
above zero: detrimental; below zero: beneficial

OSSEs+EFSO show which instruments are detrimental, when and where!

4. Value of OSSEs (B. Hooke, 1/3/2019)

Congress is right in focusing attention on OSSE's as these 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. That said, we're sure the Congress would agree that attention shouldn't be confined to OSSE's to the exclusion of other R&D including but not limited to 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 (need EISWG help here enumerating a few more). 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 tech transfer paths in parallel rather than in sequence or in isolation. More attention to OSSE's and development of their potential Is needed, but in a manner balanced by additional attention to other opportunities across the board.

It should also be emphasized that perhaps the greatest benefit of R&D on OSSE's 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."

5. Overview: OSSE Capabilities at NASA (Derek J. Posselt, 1/11/2019)

Background and Overview

NASA has been conducting Observing System Simulation Experiments (OSSEs) for decades. NASA's OSSEs have primarily been conducted by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center (GSFC) with the intent of determining 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 (c.f., Errico et al. 2013; Prive et al., 2013) and has also produced a global mesoscale-resolving Nature Run (Gelaro et al. 2015). 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 National Centers for Environmental Prediction (NCEP) operational forecast system. Simulation of observations from the Nature Run, as well as from the forecast model during a data assimilation cycle, is done using the Community Radiative Transfer Model (CRTM), in partnership with the Joint Center for Satellite Data Assimilation (JCSDA).

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. Traditional forecast OSSEs, such as those provided by the GMAO system and similar experiments conducted at NOAA, are most useful in determining the utility of new measurements for the first purpose (characterization of the state of the Earth system). Current and future OSSE activities can be used to assess the value of new observations within this context, and may benefit from comparison with OSSEs conducted by NOAA.

However, there are limitations to forecast OSSEs, especially for measuring the effectiveness of an observing system for answering specific science questions. In addition, there is a large amount of groundwork that must be done if a forecast OSSE is to provide an accurate measure of the information in a new measurement. In addressing both purposes for which NASA conducts OSSEs, it is useful to consider a spectrum of activities, many of which are essential for the success of any OSSE effort.

Components of an OSSE Spectrum

In designing a new observing system, there are several key considerations. We have organized these in approximate order of complexity, as well as in the order in which they must be completed in an end-to-end experiment.

1.   Sampling OSSE (Observing System Design Experiment; OSDE)

The first, and most fundamental consideration in observing system design is sampling; addressing the question of whether a set of measurements is able to see a feature of interest. Such a study should include the proposed new system, as well as the program of record (as it will appear at the time the new measurement comes into existence). Considerations include:

a)   Temporal sampling

The various components of the Earth system have characteristic evolution time scales and periodicity (e.g., diurnal, seasonal). An observing system should observe frequently enough to capture the temporal variability of a process of interest, and should sample completely enough to fully represent the modes of variability.

b)   Spatial sampling

Along with temporal variability, structures in the Earth system have inherent scales of spatial variability (horizontal and vertical). Measurements must be made at sufficient spatial resolution, and over a broad enough area, to encompass the natural spatial scales of variability.

A temporal and spatial sampling OSSE can be conducted without the use of a detailed instrument model, and without knowing (or specifying) instrument sensitivity. All that is required is (1) a nature run that encompasses the range of temporal and spatial scales and realistically represents the processes of interest, (2) a way of identifying features of interest, (3) a way to sub-sample the nature run according to various measurement geometries and sampling frequencies, and (4) a way to measure the difference between the full sample included in the nature run and the sub-sample returned by the prospective measurements. An assertion could be made that a sampling OSSE is the low bar; an observing system must be able to "see" a feature of interest at sufficient temporal and spatial frequency and resolution.

2.   Retrieval OSSE (Observing System Uncertainty Experiment; OSUE)

In addition to an assessment of sampling requirements, it is necessary to quantify the degree to which prospective measurements provide information on a geophysical quantity of interest. Specifically, it is important to quantify the information content in a measurement and also the requisite accuracy. Because NASA's observations consist almost entirely of remote sensing measurements, this amounts to assessment of the information about a geophysical quantity (e.g., the amount of liquid in clouds) contained in a set of indirect measurements (e.g., radar reflectivity or microwave brightness temperature). Statistical techniques can be used to quantify expected uncertainty in geophysical quantities using synthetic retrievals, and the results are used to determine whether measurements are of sufficient accuracy. In addition, the uncertainty estimates are integral in forecast OSSEs.

3.   Forecast OSSE

While NASA is not tasked with producing routine weather predictions, forecast OSSEs are useful for assessing the information contained in new observations, relative to all other current measurements. For example, while measurements of the thermodynamic structure of the atmosphere are crucial for understanding how weather systems interact with their environment(s), these measurements are already provided over much of the globe by existing infrared and microwave sounders. A forecast OSSE can be used to quantify the added benefit of a new set of observations, relative to those that are already being made, and as such are useful for both a cost-benefit analysis, and also for determining the degree to which new observations may improve the realism of the MERRA reanalysis.

References:

Errico, R. M., R. Yang, N. Privé, K.-S. Tai, R. Todling, M. Sienkiewicz, and J. Guo, 2013. Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office. Q. J. Roy. Meteor. Soc, 139, 1162-1178. doi: 10.1002/qj2027.

Gelaro, R., W. M. Putman, S. Pawson, C. Draper, A. Molod, P. M. Norris, L. Ott, N. Privé, O. Reale, D. Achuthavarier, M. Bosilovich, V. Buchard, W. Chao, L. Coy, L. Cullather, A. da Silva, A. Darmenov, R. M. Errico, M. Fuentes, M.-J. Kim, R. Koster, W. McCarty, J. Nattala, G. Partyka, S. Schubert, G. Vernieres, Y. Vikhliaev, and K. Wargan, 2015. Evaluation of the 7-km GEOS-5 Nature Run. NASA/TM–2014-104606, Vol. 36.

Privé, N., R. M. Errico, and K.-S. Tai, 2013. Validation of forecast skill of the Global Modeling and Assimilation Office observing system simulation experiment. Q. J. Roy. Meteor. Soc., 139, 1354-1363. doi: 10.1002/qj.2029.

6. When should we use OSSEs with caution (Fuqing Zhang, 1/12/2019)

I am a big fan of using OSSEs. We use OSSEs for testing new data assimilation methodologies, and evaluating the effect of new observations such as from radars and satellites. We also use OSSEs to assess the observation impacts and for observation targeting strategies, among many other applications.

However, I would like to caution the reliance on using OSSEs in designing the future NWP observation systems, and evaluate the economical values of the existing observing network, in particular, based on using simulated truth from one or a limited number of high-resolution "nature runs":

(1) The reliability and effectiveness 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. It is unclear that NOAA currently has the advanced data assimilation system ready to comprehensively evaluate the full NWP observing network at this point. For example, the current NCEP operational data assimilation system (4DenVar) has demonstrated no quantifiable impacts of any all-sky(cloudy/rainy) satellite radiances in the operational model performance (based on the very recent NCEP presentations) while the much more advanced ECMWF data assimilation and prediction system now puts all-sky radiance (minus clear-sky) as the most impactful sources of observations (see attached slide 6 from Massimo Bonivita from ECMWF presented at the AMS meeting last week). Unfortunately, the NCEP data assimilation system both in terms of DA methodology and data used is at least 5 years behind, and likely more.

(2) In other words, if you use the NCEP 4DenVar to evaluate any all-sky radiance observations, you might conclude that there will be no or little impacts for cloudy/rainy radiances but that is clearly not the case if you use the ECMWF system for the same observations in OSSE setup. Relatedly, some of you might see other talks by ECMWF at the meeting showing that as the assimilation of all-sky radiance improves, the impacts of satellite derived atmospheric motion wind decreases several folds since arguably if all the clouds and water vapor are assimilated into analysis in the right place, the impact/need for the retrieved motion wind will become less and less obvious. We are evaluating the observation network likely decades ahead, and if the not-so state-of-the-science data assimilation system is used, at the future resolution of the future modeling system, we may be making wrong choice of observations or making wrong investments.

(3) I would also caution the use of extreme events or nature runs as truth for OSSEs as well. In operational NWP, the skill scores for any model are judged by a large number of cases or seasons, and by many metrics, not a single event since individual events either 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 still significant errors in the model physics or in the boundary conditions or forcings.

Again, I am not against the use of OSSEs, and we use OSSEs a lot in our own research. However, I am cautioning the over reliance on using OSSEs to make important observation systems decision, in particular, given the apparent less-than-optimum data assimilation systems currently used for conducting such OSSEs.

Comments from Raghu: I agree with Fuqing but the contradiction would be that the same systems are used for forecasting and we want people to believe and use forecasts. So we need to tread carefully in saying that the systems are good enough for forecasts but not for OSSEs.

Other discussions:
Bill question: I think of OSSE's as focused primarily on teasing out 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 it would be possible to extend the idea to societal impacts, whether monetizable, or in terms of lives at stake, etc. Is this also routinely done? Is it a possible additional avenue of research?

Raghu answer: in an Earth System model where human component is included, one can imagine collecting human data or propagating just the physical earth system information through the human system as well.

7. **Observation System Simulation Experiments at the Naval Research Laboratory (Dan Tyndall, 1/18/2019)**

Daniel Tyndall[1], Nancy Baker[1], David Flagg[1], Charlie Barron[2], Matt Carrier[2], Scott Smith[2], Doug Allen[3], and Karl Hoppel[3]
1Marine Meteorology Division, Monterey, CA
2Oceanography Division, Stennis Space Center, MS
3Space Science Division, Washington, DC

The U.S. Navy requires meteorological and oceanographic (METOC) information to characterize the environment to support global, regional, and tactical scale operations on time scales ranging from minutes to weeks. This METOC information is provided through environmental observations collected from a wide variety of systems, processed using global and regional numerical prediction (NWP) models, and exploited using tactical decision aids (TDAs). 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 TDAs before fully investing in the systems. These estimates can be computed using an Observing System Simulation Experiment (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.

Holt et al. (2018) employed a variant of the OSSE methodology to estimate the sensitivity of Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS; Hodur 1997) forecasts to METOC sensors deployed on Navy ships and aircraft. METOC sensors aboard Navy assets were simulated using the

Global Forecast System (GFS) analysis fields; these observations were assimilated into the Navy Atmospheric Variational Data Assimilation System (NAVDAS; Daley and Barker 2001), along with other observations from the historical record. The key difference between the traditional OSSE methodology and the variant used in this study is that instead of using a nature run to simulate all observations (including observations currently part of the global observing system), only the new observations whose impacts were to be estimated were simulated using an alternate model (in this case, the GFS). The methodology used here allowed the investigation to use case studies from the historical record, instead of having to use an analogous weather situation that was never part of the historical record. Forecasts that assimilated the new observations were compared to forecasts without the new METOC sensors; this study only examined these differences in terms of sensitivity instead of absolute truth.

Simulated oceanic observations have also been used to determine potential impacts of new observations on forecasts produced by the Navy Coastal Ocean Model (NCOM; Martin 2000, Barron 2006). Helber et al. (2010a, 2010b) conducted OSSEs on the impact of observing systems based on varying numbers and configurations of unmanned underwater vehicles and satellite altimeters. Carrier et al. (2018) simulated satellite altimeter observations generated from NCOM fields to evaluate a method of assimilating direct sea surface height (SSH) observations into a free-surface model to test for reduced gravity wave contamination. These simulated observations have been created to mimic observations from the upcoming Surface Water and Ocean Topography (SWOT) sensor, which will be launched in 2021. Unlike current satellite altimetry observations, the SWOT sensor measures SSH height along a broad swath instead only at nadir. This research varied the typical OSSE methodology by simulating the SSH observations using NCOM model data from a different year (but the same month and day). A free running forecast, a forecast assimilating simulated nadir altimeter observations, and a forecast assimilating simulated SWOT observations were made using the Navy Coupled Ocean Data Assimilation (NCODA) Four-Dimensional Variational (4DVar) system. These experiments were compared to each other in order to determine the potential impact of SWOT observations on the ocean model forecast of sea surface height.

A similar methodology was employed by Smith et al. (2018) to determine the impact that potential satellite-derived salinity measurements have on NCOM forecasts. Salinity is one of the primary ocean variables affecting the dynamics of the ocean and there are currently very few of these observations that make it into the operational data stream. This OSSE generated simulated salinity observations were from a 1-km resolution NCOM nature run with an initial condition from a different year (but from the same month and day to reduce any seasonality bias) along the ground tracks of existing VIIRS satellites. OSSEs were performed with and without these salinity observations (in addition to the traditional SSH, SST and in situ observations) using the NCODA 3DVariational (3DVar; Cummings 2005) and 4DVar systems. These four OSSEs were compared with a NCOM free-running forecast and the truth to show the impact a high resolution sea surface salinity dataset would have on the Navy's 3DVar and 4DVar ocean assimilation/prediction systems.

OSSEs have also been used by NRL to study impacts of potential observations on middle atmosphere prediction. In a study by Allen et al. (2018), simulated stratospheric ozone observations and radiances were assimilated into the Navy Global Environmental Model (NAVGEM; Hogan et al. 2014). Simulated observations were computed from cycling NAVGEM analyses produced by its hybrid 4DVar data assimilation system. These analyses that were used to generate the simulated observations were created by only assimilating in-situ observations of the global observing system at pressures greater than 100 mb. Experiments compared forecasts with the simulated ozone observations to a baseline in which stratospheric initial conditions were perturbed, but did not assimilate any stratospheric observation). A separate forecast experiment was run in which forecasts with simulated ozone and radiance observations were compared to a baseline with perturbed initial conditions which only assimilated the simulated radiance observations. Another study by Hoppel (personal communication) examined the potential of a

constellation of solar-occultation (SO) sensors. In this study, SO observations were generated by subsampling NASA Aura microwave limb sounder data and were tested with NAVGEM.

NRL has started work on evaluating the historical OSSE approach as a method of producing quantitative estimates of observation impact for future and undeployed atmospheric observations. This methodology follows the approach used to determine COAMPS sensitivity from simulated METOC sensors aboard Navy ships and aircraft mentioned previously; however, this research is utilizing a high resolution coupled atmosphere and ocean provided by the Weather Research and Forecasting (WRF; Skamarock et al. 2008) model and the Regional Ocean Modeling System (ROMS; Song and Haidvogel 1994) through the Coupled-Ocean-Atmosphere-Wave-Sediment Transport Modeling System (COAWST; Warner et al. 2010). The key advantage of this approach over the traditional methodology is that the historical OSSE approach allows for the examination of actual weather events in the historical record, as well as reducing efforts required to perform the methodology as the existing observing system does not need to be simulated (it is provided from the historical record). The differences between this approach and the traditional approach are depicted in Figure 1. The historical OSSE approach is being evaluated by comparing observation impacts of real observing platforms to their simulated equivalents using cycling short-term forecast data from the COAWST reference model. This research is utilizing COAMPS and COAMPS-4DVar as the target model and data assimilation system for evaluating impacts of new atmospheric observations, which include UAV, Saildrone, and Sensor Hosting Autonomous Remote Crafts (SHARC) observations. Impacts of new ocean observations, which include acoustic Doppler current profilers, acoustic pressure observations, HF Radar velocity data, and trajectories from ocean floats and drifters will also be evaluated using historical runs of HYCOM and NCODA-3DVar as the reference and NCOM with NCODA-4DVar as the target forest model and data assimilation system.
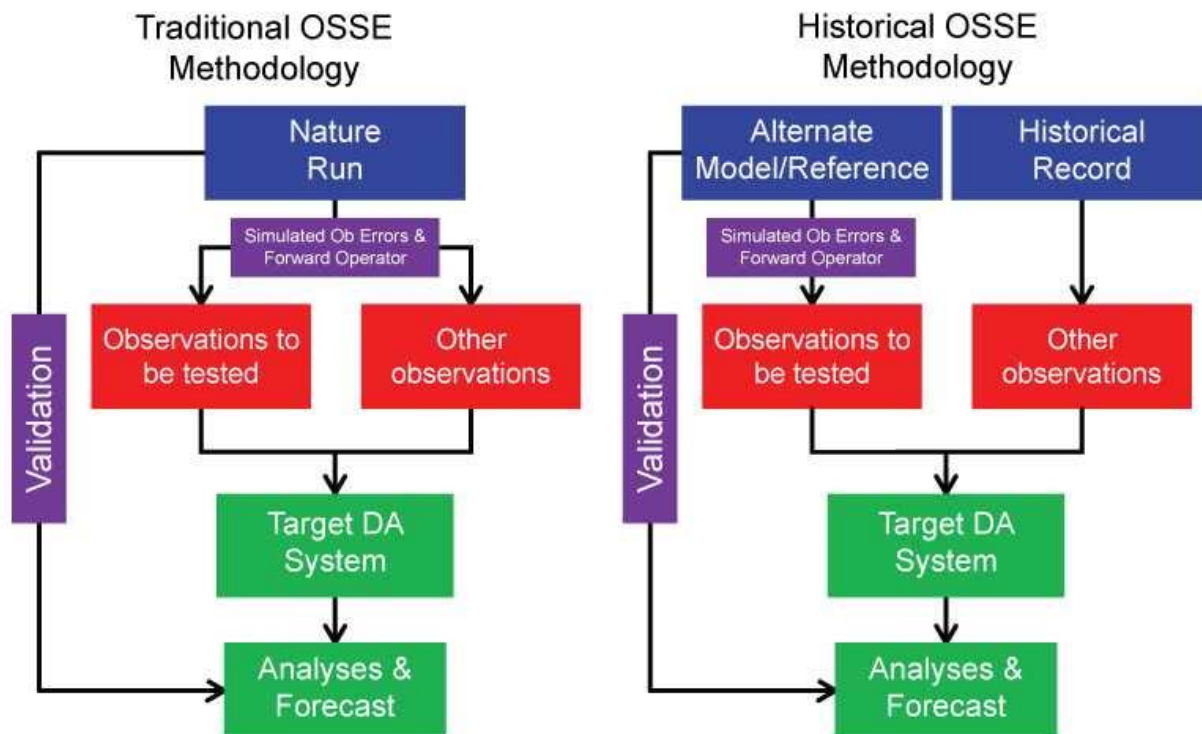

Figure 1. Schematic depiction of traditional OSSE methodology (left) and the historical OSSE methodology (right)

References

Allen, D., K. Hoppel, and D. Kuhl, 2018: Extraction of wind and temperature information from hybrid 4D-Var assimilation of strospheric ozone using NAVGEM. Atmos. Chem. Phys., 18, 2999-3026.

Barron, C.N., A.B. Kara, P.J. Martin, R.C. Rhodes, and L.F. Smedstad, 2006: Formulation, implementation and examination of vert cal coordinate choices in the global Navy Coastal Ocean Model (NCOM). Ocean Modelling, 11(3-4), 347-375, https://doi.org 10.1016/j.ocemod.2005.01.004.

Carrier, M., H. Ngodock, I. Souopgui, J. D'Addezio, S. Smith, and J. Osborne, 2018: Assimilation of satellite altimeter observations in a free-surface ocean model using a 4DVAR analysis system. 23rd Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, Phoenix, AZ, American Meteorological Society, 8.4.

Cummings, J., 2005: Operational multivariate ocean data assimilation. Quart. J. Roy. Meteor. Soc., 131, 3583–3604, https:// doi.org/10.1256/qj.05.105.

Daley, R., and E. Barker, 2001: NAVDAS: Formulation and diagnostics. Mon. Wea. Rev., 129, 869–883, https://doi.org 10.1175/1520-0493(2001)129,0869:NFAD.2.0.CO;2.

Helber, R.W., C.N. Barron, M. Gunduz, P.L. Spence and R.A. Zingarelli, 2010: Acoustic metric assessment for UUV observation system simulation experiments. U.S. Navy Journal of Underwater Acoustics, 60, 101-124.

Helber, R.W., J.F. Shriver, C.N. Barron, and O.M. Smedstad, 2010: Evaluating the impact of the number of satellite altimeters used in an assimilative ocean prediction system. Journal of Atmospheric and Oceanic Technology, 27, 528-546.

Hodur, R. M., 1997: The Naval Research Laboratory's Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). Mon. Wea. Rev., 125, 1414–1430, https://doi.org/10.1175/15200493(1997)125,1414:TNRLSC.2.0.CO;2.

Hogan, T., M. Liu, J. Ridout, M. Peng, T. Whitcomb, B. Ruston, C. Reynolds, S. Eckermann, J. Moskiatis, N. Baker, J. McCormack, K. Viner, J. McLay, M. Flatau, L. Xu, C. Chen, and S. Chang, 2014: The Navy Global Environmental Model, Oceanography, 27, 116-125.

Holt, T., J. Cummings, D. Flagg, T. Haack, B. Hughton, D. Tyndall, and A. Westgate, 2018: Electromagnetic characterization of the atmosphere in support of ASuW, NRL Memo. Report, Naval Research Laboratory, Marine Meteorology Division, Monterey, CA.

Martin, P.J., 2000: A description of the Navy Coastal Ocean Model Version 1.0. NRL Rep. NRL/FR/7322-00-9962, 42 pp.

Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G Duda, X.-Y. Huang, W. Wang, and J. G. Powers, 2008: A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp. doi:10.5065/D68S4MVH.

Smith, S., R. Schaefer, M. Carrier, J. D'Addezio, H. Ngodock, I. Souopgui, J. Osborne, and G. Jacobs, 2018: The impact of assimilating sea surface salinity observations in a high resolution ocean prediction system. 23rd Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, Phoenix, AZ, American Meteorological Society, 5.4.

Song, Y. and D. Haidvogel, 1994: A semi-implicit ocean circulation model using a generalized topography-following coordinate system. J. Comp. Phys., 115, 228-244.

Tyndall, D., D. Hodyss, N. Baker, C. Amerault, and J. Nachamkin, 2018: An examination of the historical observing system simulation experiment methodology. 23rd Conference on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, Phoenix, AZ, American Meteorological Society, 143.

Warner, J.C., Armstrong, B., He, R., Zambon, J.B., 2010. Development of a coupled ocean–atmosphere–wave–sediment transport (COAWST) modeling system. Ocean Model., 35 (3), 230–244.
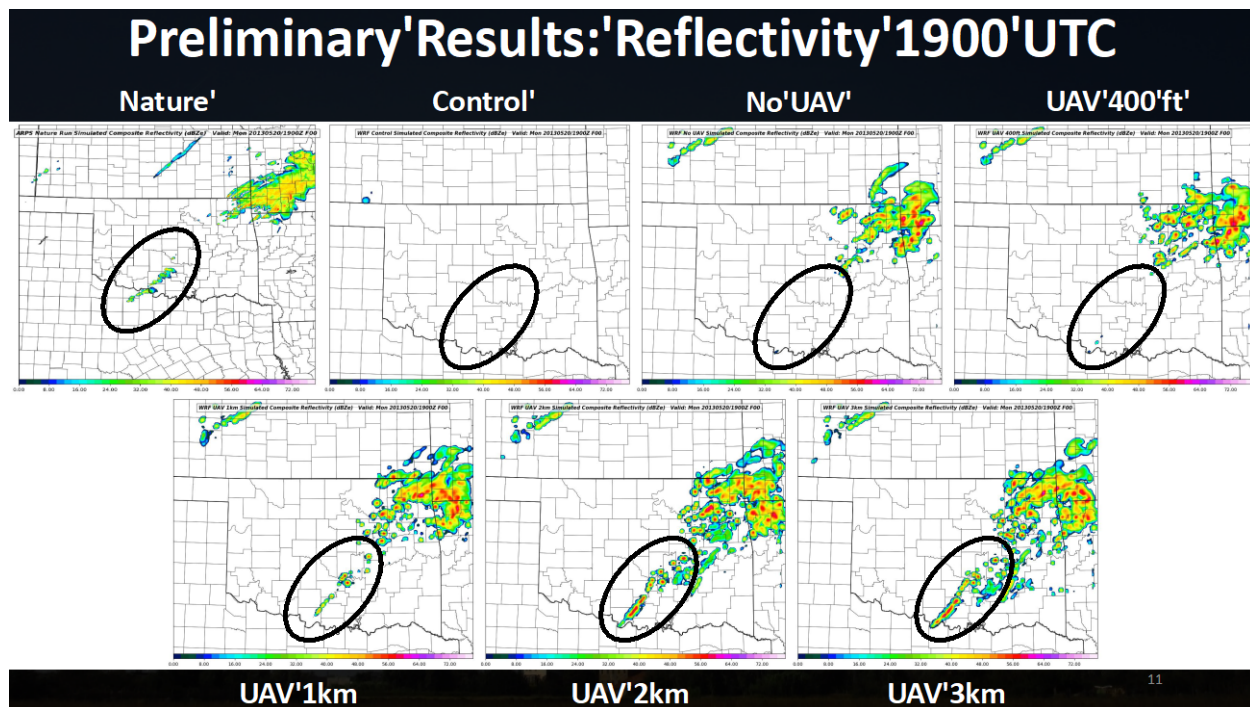
8. Assessing Impact of Observations on NWP - Summary for EISWG Report (F. Carr, 2/1/2019; Original Presentation July 26, 2018)

Observing System Experiments (OSEs)

OSEs, or "data denial experiments", are used to test value of existing observing systems to NWP (as opposed to OSSEs, which test systems not yet deployed). They work best if start with well-observed regions, so that one might also identify redundant observations. They can be done on all scales, but attention here is concentrated on mesoscale and convective-scale OSEs. Mesoscale studies by Benjamin et al at GSD examined relative value of raobs, aircraft data (ACARS, AMDAR and TAMDAR), wind profilers, GPS-IPW, VAD profiles, cloud-motion vectors and surface data over CONUS. Convective or storm-scale studies, such as those done by CAPS at OU, can assess value of radar data (Vr, Z, dual-pol variables), phased-array rapid scanning or different radar network configurations (e.g., CASA X-band "gap-filling" radars). CAPS has also assessed the value of different private sector surface networks as part of the National Mesonet Program. OSEs (as well as OSSEs) can be used to make network decisions on spatial resolution and sampling frequency, siting and scanning strategies. Several NAS 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.

Observing System Simulation Experiments (OSSEs)

While most OSSEs are done for global satellite systems, they can be used to assess new observing systems for regional scales as well. On the storm-scale, OSSEs have been done by CAPS to assess value of using dual-polarization radar data in NWP, as well as to assess different scanning strategies and network configurations. On the CONUS scale, OSSEs can be used to assess the value of increasing the density of vertical profiling systems over the U.S., as recommended by the NAS "Network of Networks" study. One recent example of such a study done by CAPS is to examine the "3-D Mesonet" concept, in which unmanned aerial vechicles (UAV, in this case, quadcopters) provide vertical soundings of temperature, wind and moisture every hour from Oklahoma Mesonet stations. A preliminary result (next page) shows that a WRF model run at 4 km is unable to simulate the beginning of a squall line seen in the Nature run with the complete conventional data set (No UAV panel), nor with UAV data up to only 400 m, but does do so when the sampling is carried out to 1 km or higher.

Preliminary Results: Reflectivity 1900 UTC

Nature / Control / No UAV / UAV 400 ft / UAV 1km / UAV 2km / UAV 3km

Forecast Sensitivity to Observations (FSOs)

Operational or research data assimilation systems 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, first done with adjoint data assimilations systems by Langland and Baker (2004) and Cardinali (2009). There is no need for data denial experiments – the operational system is the control experiment - and all observations are assessed simultaneously. It is based on adjoint sensitivity, via the tangent linear model, and is thus limited to short-range forecast (1-2 days) impacts. FSOs are useful for diagnosing short-range forecast failures due to the observations (could be a problem in proper use of the observations or due to errors – biases – in the model).

FSOs can also be done using Ensemble Kalman Filter data assimilation systems (Kalnay et al 2012), known as EFSOs. EFSOs uses ensemble perturbations to calculate the impact observations have on a forecast, without need for data-denial experiments or an adjoint. Because of the sampling issues associated with small ensembles, proper localization methods need to be developed.

Final Thoughts

• OSSE, OSE, FSO, EFSO and related methods each have their pros and cons, and should all be used to assess the relative benefit of different observing systems.
• OSSE, OSE, FSO, EFSO research efforts should be coordinated nationally to avoid duplication of effort; sharing of software tools, etc. (via QOSAP program?)
• NOAA's NOSIA-II observation system assessment tool should also incorporate OSSE, OSE, FSO, EFSO results.
• A "gap analysis" process in NOAA seems to be lacking (that is, what are the greatest new observational needs? What combination of old and new systems will work best?)
• On storm-scale, frequent, hi-res. vertical profiles of wind, temp., moisture in PBL seem to be most needed but more testing is needed.

- By the end of 2019, the results of a survey by Fred Carr of 2000 presenters at the 2017 AMS Annual Meeting on their highest priority observing needs should be available.

9. **Observing System Simulation Experiments (OSSEs) in NOAA (Robert Atlas and Lidia Cucurull, 2/5/2019)**

Observing System Simulation Experiments (OSSEs), when performed correctly, are an important tool for evaluating the potential impact of proposed new observing systems, as well as for evaluating trade-offs in observing system design, and in developing and assessing improved methodologies for assimilating new observations (Fig.1). The current methodology used for rigorous OSSEs was proposed by R. Atlas in the early 1980's, and was accepted at an international workshop in 1983 and later at the World Meteorological Organization (WMO) and the U.S. Weather Research Program (USWRP) 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.

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), forecast sensitivity observation impact (FSOI/EFSOI), and OSSEs as effective techniques to evaluate the impact of the different observation types. These studies help NOAA management prioritize mission designs in a cost-effective way by analyzing tradeoffs in the design of proposed observing systems. 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) increase NOAA's capacity to conduct quantitative observing system assessments, (2) develop and use appropriate quantitative assessment methodologies, and (3) 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 numerical weather prediction, a state-of-the-art global OSSE system based on the NASA Cubed Sphere at 7 km resolution nature run 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 we developed a 2-km state-of-the-art basin scale nature run.

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. The OSSEs included the evaluation of CYGNSS, Geo-HSS, GNSS-RO (COSMIC-2), OAWL doppler wind lidar, Cubesats (MicroMas and CIRAS), UAS with Global Hawk and G-IV aircrafts, and targeted dropsondes for Pacific mid latitude winter storms. OSEs evaluated the impact of wind lidar onboard the P3 aircraft, GNSS-RO (COSMIC-1), data gap mitigation strategies, and advanced assimilation algorithms for GNSS-RO and GOES-R lightning data. Additionally, we conducted OSSEs related to the role of ocean observations in hurricane

prediction. 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.
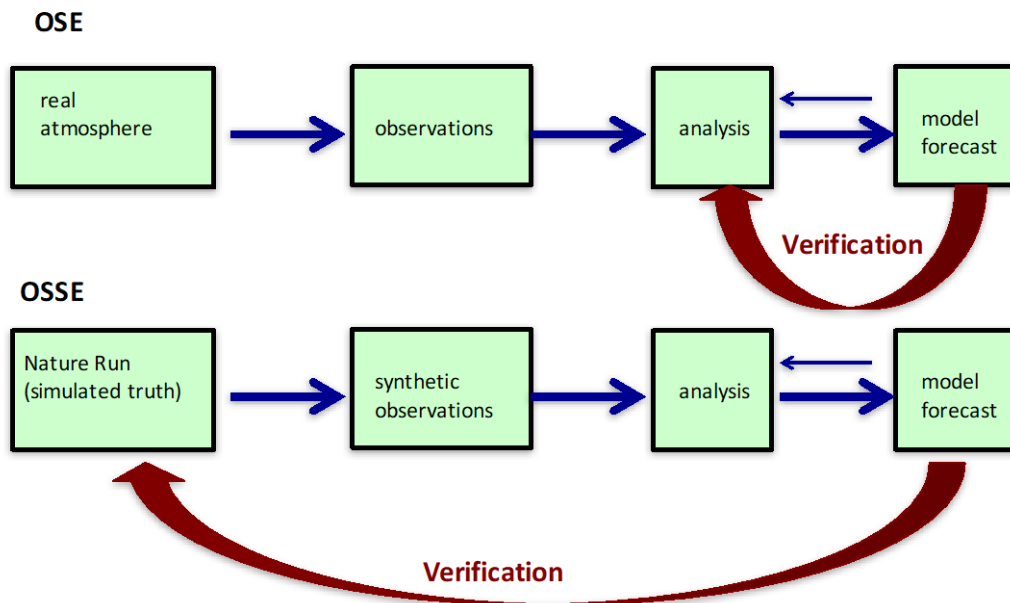

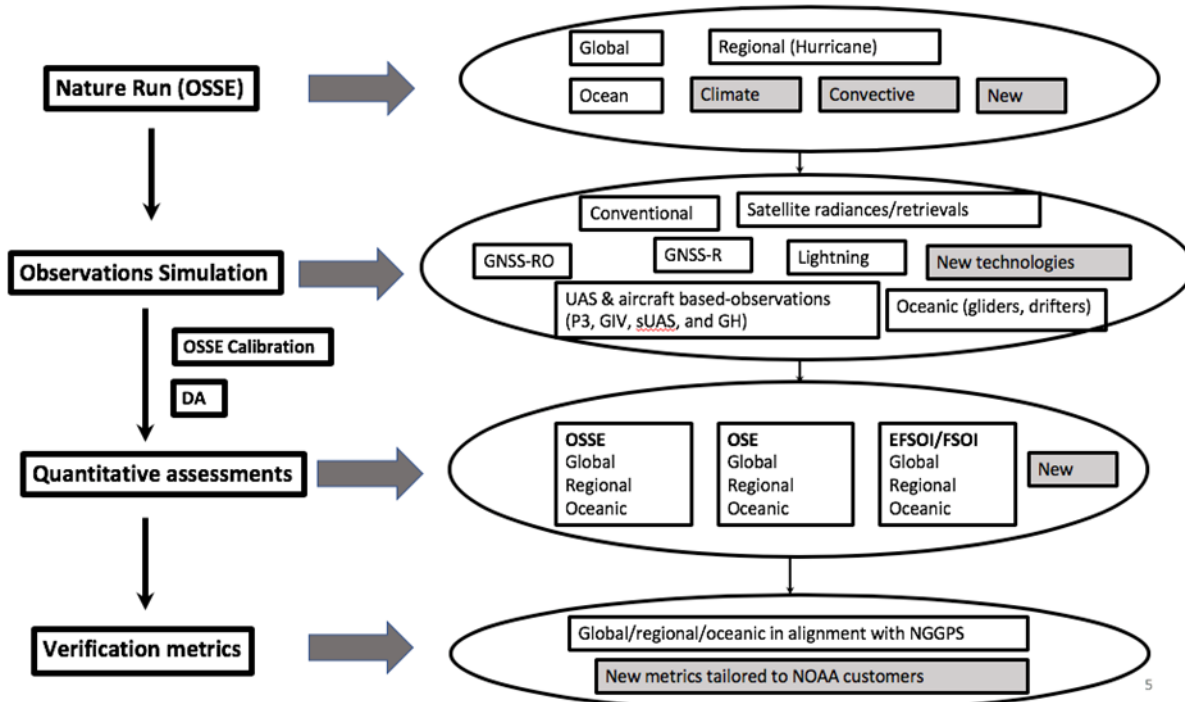Fig.1. Cartoon showing the differences between OSE and OSSE methodologies.

Fig.2. Diagram showing the different components of the OSSE system at AOML.

10. **Some thoughts on discussion and findings (EISWG Co-Chair Brad Colman, presented to the NOAA SAB on 11/1/2018)**
- OSSEs have a lot of potential, but cannot separate the impact of the different observing systems on the forecasts.
- Ensemble Forecast Sensitivity to Observations (EFSO) can evaluate during the 6hr forecasts whether each observation is beneficial or detrimental.
- Combining OSSEs with EFSO (Chen and Kalnay, 2018) will provide much more information about each observing system as shown in the next two slides.
- This will make OSSE+EFSO much more effective and useful (showing which instruments are detrimental, when and where).

11. **Email messages (Xubin Zeng' email on 1/12/2019 after the face-to-face meeting during the AMS Annual Meeting in Phoenix)**

On the last bullet (recommendations to NOAA), a few big ideas were mentioned at the meeting:

- use OSSEs 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)
- accelerate the OSSE development for both weather and earth system models (e.g., for sea ice prediction)
- develop global 5 km (and preferably 3 km) Nature Run based on earth system models as the basis for a variety of OSSEs. This may require the purchase of new high-performance computers.
- Use OSSEs to assist the exploration of strategies for the most effective and efficient way to do sea ice prediction (observations, models, DA). Should NOAA request ice-breakers? How many?
- Use OSSEs to compare the value of planned (small number of large) satellites versus (large number of small and cube) satellites for weather and climate prediction.

Additional interesting points from the discussions:

- Besides full-scale OSSE experiments, simple experiments could be very powerful (e.g., for sampling strategies and data value evaluation).
- 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.