Uncertainty of Observation Impact Estimation in an Adjoint Model Investigated with an Observing System Simulation Experiment

N. C. PRIVÉ AND R. M. ERRICO
Goddard Earth Sciences Technology and Research Center, Morgan State University, Greenbelt, Maryland

(Manuscript received 2 April 2019, in final form 7 June 2019)

ABSTRACT

Adjoint models are often used to estimate the impact of different observations on short-term forecast skill. A common difficulty with the evaluation of short-term forecast quality is the choice of verification fields. The use of self-analysis fields for verification is typical but incestuous, and it introduces uncertainty resulting from biases and errors in the analysis field. In this study, an observing system simulation experiment (OSSE) is used to explore the uncertainty in adjoint model estimations of observation impact. The availability of the true state for verification in the OSSE framework in the form of the nature run allows calculation of the observation impact without the uncertainties present in self-analysis verification. These impact estimates are compared with estimates calculated using self-analysis verification. The Global Earth Observing System, version 5 (GEOS-5), forecast model with the Gridpoint Statistical Interpolation system is used with the National Aeronautics and Space Administration Global Modeling and Assimilation Office (NASA/GMAO) OSSE capability. The adjoint model includes moist processes, with total wet energy selected as the norm for evaluation of observation impacts. The results show that there are measurable but small errors in the adjoint model estimation of observation impact as a result of self-analysis verification. In general, observations of temperature and winds tend to have overestimated impacts with self-analysis verification while observations of humidity and moisture-affected observations tend to have underestimated impacts. The small magnitude of the differences in impact estimates supports the robustness of the adjoint method of estimating observation impacts.

1. Introduction

Modern operational numerical weather prediction (NWP) relies on the ingestion of tens of millions of observational data every day in order to produce the best possible analyses and forecasts. It is of interest to be able to determine how these data influence the quality and skill of the NWP forecasts when designing a new observing system, evaluating the performance of data assimilation systems (DAS), or monitoring operations for possible issues. Forecast sensitivity to observation impact (FSOI) methods include a variety of different techniques for estimating how observations ingested by data assimilation systems affect the quality of forecasts in NWP.

One FSOI method uses an adjoint model that is based on linearized relationships between observations, analyses, and forecast increments (Langland and Baker 2004). The adjoint model can be used to estimate the impact of each ingested observation on a selected metric. In general, some measurement of forecast error norm is selected as this metric. Because adjoint models are based on the assumption of linearity, they are generally assumed to be valid only for relatively short forecasts of perhaps 24–48 h maximum for the atmosphere. This introduces a complication of how to verify the error norm; in most cases, the analysis from the full model is used for verification. However, the shorter the forecast time, the more incestuous the verifying analysis field is relative to the total forecast error (Privé and Errico 2015).

Daescu (2009) explored the effects of uncertainty of the verification fields on adjoint estimation of observation impacts using the Lorenz 40-variable model. He showed that it is possible for the adjoint estimation of observation impact to be of incorrect sign (i.e., assigning a negative impact when the true impact is positive or vice versa) under some circumstances, with less confidence in the impact estimation when the data assimilation system has imperfect error statistics. Necker et al. (2018) has addressed the similar issue of the choice of verification method for FSOI using ensemble techniques rather than...
an adjoint model, and found that model and observation biases could substantially affect the calculated impacts.

**Todling (2013)** and **Cardinali (2018)** used observation-based verification techniques to explore the question of self-verification uncertainty in adjoint estimates of observation impact. Because the observation-based metrics are in observation space, their results cannot be directly compared to the state space metrics from the adjoint. **Todling (2013)** found that self-analysis verification led to a modest overestimation of beneficial observation impacts on the 24-h forecast skill for most data types. **Cardinali (2018)** had more varied findings, with some data types having significantly underestimated beneficial impacts with self-analysis verification methods and other data types having overestimated beneficial impacts.

Observing system simulation experiments (OSSEs) are a unique tool that can be used to examine the influence of verification quality on adjoint estimations of observation impacts. An OSSE is a pure modeling experiment, wherein the real world is replaced by a nature run (NR) that is treated as the “truth” so that the true verification state is available. Simulated observations and their errors are generated from the NR for all observing types typically used in operational numerical weather prediction. These synthetic observations are then ingested using the data assimilation system and forecast model of choice, and used to make analyses and forecasts that can be compared with the NR to calculate errors that are not calculable in the real world.

An adjoint can also be used in the OSSE framework, where additional options for verification are available. The adjoint may be performed using the self-analysis state as verification of a short-term forecast, as is the usual procedure with real data. Self-analysis verification can be used to validate the performance of the OSSE system in comparison to real observations. The NR can instead be used as verification for the adjoint, however, allowing for adjoint sensitivities to be derived for very short forecasts.

This paper attempts to address some of the questions that have been raised regarding the role of verification uncertainty in adjoint estimation of observation impact using an OSSE. The primary question is whether the uncertainty of self-analysis verification substantially affects the adjoint estimation of observation impacts, both in the sense of overall statistics for a data type and for individual observations. The spatial distribution of uncertainty in short-term forecast errors and adjoint estimation of observation impacts is also explored.

The experimental setup and methods are described in section 2. Validation of the OSSE framework and adjoint is described in section 3. Experimental results are presented in section 4, and conclusions are discussed in section 5.

**2. Method**

The Global Modeling and Assimilation Office at the National Aeronautics and Space Administration (NASA/GMAO) has developed an OSSE framework using the Gridpoint Statistical Interpolation (GSI) data assimilation system (Kleist et al. 2009) and the Global Earth Observing System, version 5 (GEOS-5), NWP model (Rienecker et al. 2008). For these experiments, the three-dimensional variational analysis version of the GSI was used. The GMAO OSSE performance and validation are described in detail in Errico et al. (2013) and Privé et al. (2013) for an earlier version of the system; a synopsis of the current OSSE framework is given here.

The GMAO OSSE uses simulated conventional and radiance observations generated from a nature run. The NR is a 24-month free-running forecast of the GEOS-5 model at 7-km horizontal resolution with 72 vertical levels (Gelaro et al. 2014), and output saved every 30 min. Although the GEOS-5 model is used for both the NR and the experimental OSSE forecasts, there are some differences in the model versions and resolutions, so that an identical twin case is avoided. Specifically, the experiments are run at approximately 25-km resolution instead of the 7-km resolution of the NR, and the NR was created using the version of the GEOS-5 model from 2014, while the OSSE forecasts use a version developed in 2018. A more sophisticated two-moment microphysics scheme (Barahona et al. 2014) is employed in the GEOS-5 forecasts for the experiments, while a single-moment microphysics was used for the NR. Some other differences in the model physics include the choice of boundary layer and low-level cloud parameterizations. Despite these changes to the model physics, the OSSE framework can be considered to be a “fraternal twin” case in which the forecast model has substantially less model error in comparison to the NR than real world model error.

The simulated or synthetic observation types include most of the data types ingested in operational numerical weather prediction during 2015. Conventional types include rawinsondes, aircraft observation types, GPS radio occultation (GPS-RO) bending angles, atmospheric wind vectors from infrared, visible, and water vapor fields, surface data types, and scatterometers. Synthetic radiance observations include CrIS, IASI, AIRS, HIRS-4, SSM/IS, ATMS, MHS, and AMSU-A microwave and infrared data. Radiance observations are generated using the Community Radiative Transfer Model (CRTM;
Han et al. (2006), while conventional types are generated through spatiotemporal interpolation of the NR fields. GPS-RO bending angles are generated using the operator developed by Culverwell et al. (2015). Simulated errors are added to the synthetic observations in order to mimic the statistical error characteristics of real data, including both uncorrelated errors, and errors that may be correlated horizontally, vertically, or between channels, depending on the observation type. Full details of the synthetic observation and simulated error generation can be found in Errico et al. (2017).

An adjoint of the GEOS-5 model has been developed that includes moist physics (Holdaway et al. 2014) and is regularly employed to estimate observation impacts. The validation of the GMAO OSSE framework in terms of the adjoint observation impacts is discussed in section 3.

The observation impact can be evaluated with a combination of complementary metrics, including the net impact, the impact per observation, and a fractional impact. The net impact is defined here as the total impact of all observations of each data type, averaged over 2 months of once-daily forecasts, where a negative impact indicates a reduction in error of the 24-h forecast. The impact is calculated as a mean daily value over the 2-month period.

A comparison of the net global impacts for the Real and OSSEANA cases are shown in Fig. 1. Error bars in all figures indicate the 95% confidence interval, calculated with the daily mean impacts for each observation type considered as one data point rather than individual observations. In general, it is expected that the OSSEANA will have smaller impacts than the Real case, because of insufficient model error in the OSSE. In a cycling NWP system where the background and analysis error statistics are relatively stable with time, the “work” done by the observations is balanced by the growth of errors between cycle times. The lack of model error growth results in less “work” for the observations to perform. Indeed, for most observation types, OSSEANA displays substantially (30%-60%) less impact than the Real case, with exceptions for atmospheric wind vectors, aircraft temperatures, and some surface data types.

Figure 2 shows the fractional global observation impact for the Real and OSSEANA cases. The fractional impact is the net impact for each observation type normalized by the total impact of all observing types, giving the fraction of total impact for each data type:

$$F_j = I_j / \sum_k I_k,$$

where $F_j$ is the fractional impact of the $j$th data type, and $I_k$ is the time mean impact of the $k$th data type, and there are $k$ types of data types ingested. Even if the OSSE is expected to have smaller impacts than the Real case, ideally the two
cases should have similar fractional impacts for each data type so that the relative importance of observations is replicated in the OSSE framework. For ATMS, AMSU-A, SSM/IS, GPS-RO, rawinsonde humidity and temperatures, and aircraft winds and temperatures, the fractional impacts are very similar in the Real and OSSE cases. For rawinsonde winds, IASI, AIRS, and CRIS, the fractional impact in OSSEANA is lower than in the Real case, and for atmospheric wind vectors (AMVs) and MHS, the fractional impact in the OSSE is higher. It is possible that the lower impacts for the hyperspectral instruments may be due to difficulties in assigning correlated errors. Although there are some substantial differences in fractional impact between the Real and OSSEANA cases, the relative ranking of the importance of the various data types is generally well preserved, with AMVs being a notable exception.

The observation impact can also be calculated on a per-observation basis, where the net impact of each data type is divided by the number of observations of that type. The per-observation impact normalizes any differences in the number of ingested observations that may occur in the two cases. As with the net observation impact, it is expected that per-observation impacts will be somewhat smaller in the OSSEANA case than in the Real case because of insufficient model error. This is seen for all radiance types, aircraft winds, rawinsonde temperatures, and GPS-RO (Fig. 3), but rawinsonde wind and humidity, surface wind, and atmospheric motion vector observations have similar per-observation impacts in the Real and OSSEANA cases. Rawinsondes have more data points in the Real case than in the OSSEANA case, because of the large number of vertical levels reported in real rawinsonde soundings. (In the OSSE, the limited number of model levels in the G5NR leads to a reduced number of reported vertical levels in each sounding.) The larger than expected fractional impact for atmospheric motion vectors may in part be due to compensation for the lower observation count of rawinsonde winds.
For the purposes of this study, it is not imperative that the OSSE perfectly replicate the observation impacts seen with real data but merely that the OSSE be able to produce reasonable estimates of observation impacts. Although the net impacts are substantially weaker in OSSEANA, and some of the data types have relative impacts that are too high or too low, the overall performance of the FSOI in the OSSE framework is judged to be sufficiently robust for further studies.

4. Experiments

Adjoint estimations of observation impacts are calculated for the OSSE case using two different verification datasets. In the OSSEANA calculations, the analysis field from the OSSE cycling is used as verification, while for the OSSENR calculations, the nature-run fields are used as verification. As the forecast fields are run at C360 horizontal resolution, but the analysis fields are calculated at C180 resolution, the nature-run fields are interpolated to C180 resolution for the purposes of all FSOI calculations.

First, the analysis error is calculated over the two-month experimental period (Fig. 4) by differencing the analysis and NR fields. The zonal-mean square root temporal-mean square (RMS) temperature analysis error has maxima near the surface at both poles, and also in the lower troposphere between 0° and 30°S. Although the OSSE has a fraternal twin setup, the differences in forecast model physics between the nature run and the more recent version of GEOS-5 used for the experiment forecasts are such that there are some differences in large-scale climatology. One particularly large difference in climatology involves low-level clouds over cool ocean, such as off the western coasts of North and South America and southern Africa. This difference in cloud fraction results in a strong bias in temperatures that is not fully corrected in the analysis by the available radiance observations, resulting in a persistent region of large zonal-mean temperature analysis error in the lower troposphere between 0° and 30°S.

The highest specific humidity analysis errors are in the tropical lower troposphere, as is expected due to the dependence of $q$ on temperature in a nearly saturated
environment. The modifications to the forecast model physics compared to the NR, such as using two-moment microphysics and selecting different boundary layer parameterizations that influence convective processes, are expected to have a substantial effect on the humidity fields that are directly impacted by moist convection. The greatest zonal wind errors are in the near-equatorial region, peaking near the tropopause, with flanking local error maxima near the extratropical jets in both hemispheres.

![Figure 3](image_url)  
**FIG. 3.** Per-observation adjoint estimated daily mean observation impact on total wet energy for select data types (J kg\(^{-1}\)). Black is the Real case; gray is the OSSEANA case. Negative values indicate a reduction in the 24-h forecast error; note the direction and scale of the x axis. Error bars indicate 95% confidence intervals.

![Figure 4](image_url)  
**FIG. 4.** Zonal-mean temporal root-mean-square analysis error in the OSSE for (left) temperature (K), (center) specific humidity (kg kg\(^{-1}\)), and (right) zonal wind (m s\(^{-1}\)).
The 24-h forecast error is also of interest, as the error field is an important factor in the adjoint calculations of observation impact for the selected norm. Figure 5 shows the RMS 24-h forecast error calculated using the self-analysis verification (top), the NR verification, which yields the so-called true error (center), and the relative fraction of true error captured by the self-analysis verification (bottom) for the temperature, humidity, and zonal wind. There is clearly a substantial difference between the true and estimated forecast error for all three fields, with the true forecast error having larger magnitude. Temperature shows a particularly striking difference, with the self-analysis estimated forecast error missing some lower-tropospheric areas of high error, as well as smaller errors overall. For example, the self-analysis verified forecast temperature error excludes the regions of strong temperature bias between the NR and forecast model in the lower tropical troposphere. Because the errors in those regions are due to differences in cloud climatology that are persistent in time and are largely uncorrected in the analysis, the error regions are missed by the self-analysis verification.

The fraction of the true forecast error captured by the self-analysis verification is less than 60% in some regions, such as for temperature and winds in the tropics. The largest discrepancies in 24-h forecast error in the troposphere are found between 0° and 30°S, particularly for temperature and winds. These deep column tropical differences are related to changes in the preferred climatology of the Hadley circulation between the NR and the forecast model. The NR Hadley circulation has a tendency toward a double intertropical convergence zone region just south of the equator (Gelaro et al. 2014), and this tendency is weaker in the forecast model. This type of persistent climatological error is difficult for the DAS to correct due to the decoupling of the wind and temperature
fields in the deep tropics, so the analysis field tends to reflect the forecast model climatology rather than the NR climatology of the tropospheric tropical state.

Figure 6 compares the estimated global observation impact for the OSSEANA and OSSENR cases. For some data types, the observation impacts are not statistically significantly different between the two verification methods, but significant differences are found for rawinsonde winds, AMVs, IASI, and AMSU-A, all of which have more negative estimated impact in the OSSEANA case. These are the four observation types with the greatest net impact of all observations ingested. The relative ranking of data types by impact is in agreement between the two cases.

Fractional differences in adjoint estimations of observation impact in the OSSEANA and OSSENR calculations for the global, NHEX (20°–70°N), SHEX (20°–70°S), and tropics (20°S–20°N) regions are given in Table 1. For the global region, the impact estimation differences are positive (less improvement) in the OSSENR case compared to the OSSEANA case for most observation types, with the exceptions of MHS, surface temperatures, and...
rawinsonde humidity. In the NHEX and SHEX regions, the total impact of all observing types in OSSENR is approximately 18% smaller than for OSSEANA, but for the tropics region, the total impact in OSSENR is only 9% smaller. There are some exceptions to the general finding of smaller impacts in the OSSENR case, including MHS, surface temperature, and rawinsonde humidity in all regions and GPS-RO, SSM/IS, rawinsonde temperatures, and ATMS in the tropics.

Particularly large fractional differences in adjoint estimation are found for CrIS, HIRS4, surface pressures and temperatures, and rawinsonde humidities; however, this may be a consequence of noisiness of the estimates as all of these types have relatively small overall impacts. Small fractional differences are found for rawinsonde temperatures, MHS, ATMS, SSM/IS, and AMSU-A.

The correlations of the observation impacts calculated with the two verification methods can be calculated for each data point. These are provided for each observation type over the 2-month period of 0000 UTC analyses in Table 2. Overall, correlations are high, above 0.7 for all data types, and above 0.9 for many types. The lowest correlations are found for observations that involve humidity, such as rawinsonde specific humidity, surface humidity observations, and the MHS platforms.

The humidity field is strongly impacted by convective processes and is thus expected to have high variability at small scales. The lower correlations of observation impacts for the humidity fields are believed to be due to the noisiness of the humidity and humidity error fields.

Density plots showing the relationship between pairs of observation impacts for the two verification methods are shown for select data types in Fig. 7. The distributions of pairs in Fig. 7 are generally close to the diagonal, with the greatest number of observations having impacts close to zero for both verification methods. Humidity-based data types (rawinsonde specific humidity and MHS n18 in Fig. 7) have a somewhat more skewed distribution, with more data points having an OSSEANA impact closer to zero than the corresponding OSSENR impact.

Millions of data are ingested into the operational DAS each cycle, but Gelaro and Zhu (2009) found that on average, only slightly more than half of all global observations have a negative (i.e., beneficial) adjoint impact on the 24-h forecast. For the REAL data case, 51.3% of observations have negative impact. In the OSSEANA case, 51.0% of observations are estimated to have a beneficial impact, while in the OSSENR case 50.6% of observations have a beneficial impact. These results differ from those of Cardinali (2018), who found that the percentage of observations with negative impacts increased by 1% or more for most data types when verification of the adjoint was performed using observations rather than self-analysis. Todling (2013) compared an observation-space-based metric of forecast impact with a state-space-based energy norm, and found that the percent of beneficial observation impacts was much larger with the observation-based metric. The difference in results compared to Cardinali (2018) are likely due to the use of the different choice of metrics.

To examine the spatial distribution of impacts, for each data type the total impact difference between

---

**Table 1.** Difference in adjoint impact as a fraction of total impact (OSSENR – OSSEANA)/OSSENR for various data types in four regions; values with magnitude of 0.25 or greater are in boldface type, and negative values are in italics. Global covers 90°N–90°S, NHEX is 20°–70°N, SHEX is 20°–70°S, and the tropics are 20°N–20°S. Here, raob indicates rawinsonde observations, ps_land and ps_ocean are surface pressure observations over land and ocean, respectively, and W indicates wind observations.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Global</th>
<th>NHEX</th>
<th>Tropics</th>
<th>SHEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRS</td>
<td>0.26</td>
<td>0.27</td>
<td>−0.05</td>
<td>0.86</td>
</tr>
<tr>
<td>AMSU-A</td>
<td>0.11</td>
<td>0.07</td>
<td>0.04</td>
<td>0.16</td>
</tr>
<tr>
<td>ATMS</td>
<td>0.04</td>
<td>0.02</td>
<td>−0.10</td>
<td>0.12</td>
</tr>
<tr>
<td>CrIS</td>
<td>0.88</td>
<td>0.41</td>
<td>0.62</td>
<td>2.06</td>
</tr>
<tr>
<td>HIRS4</td>
<td>1.85</td>
<td>1.81</td>
<td>9.18</td>
<td>0.45</td>
</tr>
<tr>
<td>IASI</td>
<td>0.46</td>
<td>0.18</td>
<td>0.58</td>
<td>0.43</td>
</tr>
<tr>
<td>MHS</td>
<td>−0.10</td>
<td>−0.10</td>
<td>−0.13</td>
<td>−0.10</td>
</tr>
<tr>
<td>SSM/IS</td>
<td>0.09</td>
<td>0.15</td>
<td>−0.14</td>
<td>0.10</td>
</tr>
<tr>
<td>ps land</td>
<td>0.54</td>
<td>0.84</td>
<td>−0.44</td>
<td>1.17</td>
</tr>
<tr>
<td>ps ocean</td>
<td>−0.40</td>
<td>0.84</td>
<td>−0.28</td>
<td>0.87</td>
</tr>
<tr>
<td>Surface T</td>
<td>−0.75</td>
<td>−0.58</td>
<td>−0.48</td>
<td>−0.93</td>
</tr>
<tr>
<td>Surface q</td>
<td>0.63</td>
<td>0.61</td>
<td>0.67</td>
<td>0.52</td>
</tr>
<tr>
<td>Surface W</td>
<td>0.30</td>
<td>0.38</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>AMV</td>
<td>0.28</td>
<td>0.28</td>
<td>0.33</td>
<td>0.25</td>
</tr>
<tr>
<td>Aircraft T</td>
<td>0.33</td>
<td>0.50</td>
<td>0.11</td>
<td>0.14</td>
</tr>
<tr>
<td>Aircraft W</td>
<td>0.30</td>
<td>0.37</td>
<td>0.21</td>
<td>0.34</td>
</tr>
<tr>
<td>Scatwind</td>
<td>0.21</td>
<td>0.29</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>Raob T</td>
<td>−0.03</td>
<td>0.01</td>
<td>−0.10</td>
<td>0.06</td>
</tr>
<tr>
<td>Raob q</td>
<td>−0.52</td>
<td>−0.35</td>
<td>−0.59</td>
<td>−0.65</td>
</tr>
<tr>
<td>Raob W</td>
<td>0.40</td>
<td>0.31</td>
<td>0.58</td>
<td>0.51</td>
</tr>
<tr>
<td>GPS-RO</td>
<td>0.12</td>
<td>0.52</td>
<td>−0.27</td>
<td>0.34</td>
</tr>
</tbody>
</table>

---

**Table 2.** Correlation between adjoint estimated impacts for various observation types calculated with self-analysis verification vs nature-run verification.

<table>
<thead>
<tr>
<th>Obs</th>
<th>Correlation</th>
<th>Obs</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rawinsonde T</td>
<td>0.92</td>
<td>GPS-RO</td>
<td>0.94</td>
</tr>
<tr>
<td>Aircraft T</td>
<td>0.87</td>
<td>Surface T</td>
<td>0.88</td>
</tr>
<tr>
<td>Rawinsonde q</td>
<td>0.71</td>
<td>Surface q</td>
<td>0.74</td>
</tr>
<tr>
<td>Surface pressure</td>
<td>0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rawinsonde wind</td>
<td>0.89</td>
<td>AMV</td>
<td>0.92</td>
</tr>
<tr>
<td>Aircraft wind</td>
<td>0.85</td>
<td>Surface wind</td>
<td>0.87</td>
</tr>
<tr>
<td>AIRS Aqua</td>
<td>0.94</td>
<td>AMSU-A</td>
<td>0.94–0.95</td>
</tr>
<tr>
<td>ATMS</td>
<td>0.94</td>
<td>CrIS</td>
<td>0.91</td>
</tr>
<tr>
<td>HIRS-4</td>
<td>0.92</td>
<td>IASI</td>
<td>0.93</td>
</tr>
<tr>
<td>MHS</td>
<td>0.78–0.79</td>
<td>SSM/IS</td>
<td>0.94</td>
</tr>
</tbody>
</table>
OSSENR and OSSEANA over the 2-month period is summed in bins that are 4° by 4° in latitude and longitude, with all vertical levels (or channels for radiance data) combined in each bin, then converted to a daily average. Select maps of this summed impact difference are shown in Fig. 8, where negative values indicate that OSSEANA has either more negative or less positive values than for OSSENR (i.e., that the observations appear less beneficial or more harmful when validated using the truth). The magnitude of the impacts shown in Fig. 8 may be compared with the magnitude of the bars on the right side of Fig. 6, where the area-weighted sum of the impacts in each map will sum to the total in each bar. Conventional data types are similarly binned in vertical levels to calculate daily-average zonal-mean impacts in Fig. 9.

The results shown in Figs. 8 and 9 are in line with the net impact differences shown in Table 1. GPS-RO has a more negative net global impact in the OSSEANA case, but a more positive impact in the TROPICS region. There are localized regions of positive and negative impact differences for all observation types, some regions appear noisy in character while other regions show more coherent differences.

Fig. 7. Density plot of paired estimates of adjoint impact; the scale is log_{10} count of the number of observations per 10^{-10} J kg^{-1} (grid spacing 10^{-5} J kg^{-1} on both axes). The abscissa is the OSSEANA impact (J kg^{-1}); the ordinate is the corresponding OSSENR impact. The red line indicates x = y.
and likely represent some type of feature or persistent characteristics.

For example, AMVs have strongly negative differences over the Pacific ocean tropical region, except for a narrow positive difference band near the equator. Figure 9 indicates that the largest differences in AMV impact are in the upper troposphere near the intertropical convergence zone outflow region, where the analysis error is large (Fig. 4). Because AMVs are generally confined to altitudes below 150 hPa, the maximum difference in AMV impacts is cut off below the higher altitude of the maximum in analysis error at the equator. These upper-tropospheric AMV observations are confined to regions of deep convection where clouds are present, so there is a high density of AMVs between 0° and 20°N over the eastern and central Pacific but few AMV observations over the upper levels in the Southern Hemisphere tropics in this same region. While the GOES East and West AMVs have a high density of observations in the tropical upper troposphere, Meteosat and HIMAWARI AMVs over the western Pacific and Indian Oceans have relatively sparse upper-level AMV observations, which may be the cause of regional variability in impact differences in the tropics seen in Fig. 8.

5. Conclusions

The first order finding from these experiments is that while there are some differences in the estimated forecast error depending on the choice of verification method, the adjoint estimates of observation impact are relatively similar overall. The largest discrepancies in adjoint estimation of impacts occur for conventional wind data types, IASI, and AMSU-A in the Southern Hemisphere. Observations of temperatures and wind fields have smaller impact with the “truth” (OSSENR) verification, while observations of humidity have stronger impact. Only the data types with largest magnitude of forecast impact have statistically significant differences in estimated impact depending on verification method.

Figure 5 demonstrates that the self-analysis verification underestimates the magnitude of errors in the 24-h
forecast fields. Since the analysis field used to verify the forecast is four cycles of the DAS later than the analysis used to start the corresponding forecast, this underestimation is expected to be a result of model biases that are not fully corrected by the ingestion of observational data. Data types that have persistent bias or temporally correlated errors could also contribute to the underestimation of forecast error. Note that in this OSSE, the simulated observational errors are not temporally correlated, so that any temporal correlation in observation errors is due to intrinsic errors stemming from the observation operators used to simulate and ingest the data.

Examination of the differences in error norms (not shown) between the two verification methods found little correlation between the location of observations with large differences in impacts in the OSSEANA and OSSENR calculations and those regions with large differences in error norm. Since the error norm is calculated after the observation time, this implies that the model operator projects the observation impacts a substantial distance during the 24-h integration period.

The greatest discrepancies in analysis error and estimated forecast error are generally found in the tropics. However, the regions where individual observations have the greatest difference in estimated observation impact depending on verification method vary for different data types. There is not a clear pattern to the difference in observation impacts with the two verification methods in terms of geography.

The ratio of beneficial to detrimental observations was only slightly decreased with the use of the truth for verification field. The greatest fraction of beneficial observations is expected to occur at the analysis time, with this fraction decreasing as the impact of observations is considered for forecast times of increasing length. For a sufficiently long forecast, the total observation impact would be expected to

**Fig. 9.** Zonal-mean 2-month daily-average difference in adjoint impact estimates (J kg$^{-1}$), OSSEANA – OSSENR for select observing types: (top left) GPS-RO; (top right) atmospheric motion vectors; (bottom left) aircraft temperatures; (bottom right) rawinsonde temperatures.
asymptote to zero, and the fraction of beneficial observations would be 50% (i.e., each observation would simply randomly perturb the forecast state; Ehrendorfer 2007). As the fraction of beneficial observations in the OSSEANA case is only 51.0%, any reduction in beneficial impact in the OSSENR case compared to OSSEANA is confined to be 1.0% or less.

The correlations between FSOI for individual observations with different verification methods are high. There are few observations that have a strong impact of one sign with one verification method but an impact of different sign with the other verification method. Figure 7 indicates that most observations with opposite signs for the two verification methods (quadrants II and IV) have small or near-zero impacts. The sign and magnitude of the discrepancy in adjoint impact estimates due to verification uncertainty is highly variable in location, as seen in Fig. 8, so blanket assumptions of weaker or stronger impacts to adjust for uncertainty on a global scale are unwarranted. Daescu (2009) found that larger uncertainties in the adjoint estimate of observation impact were found when the error characteristics of the observations assumed by the data assimilation system were incorrect. This could imply that the observation types with the greatest discrepancies in impact are mishandled by the data assimilation system in some way.

The results described in this paper show mixed agreement to previous work by Todling (2013) and Cardinali (2018). Both of these previous studies used observation-based verification and metrics, which is different from using a state-space field for verification and likely accounts for much of the difference in results. In terms of observation impact estimates, Todling (2013) found that the observation-verified impacts were slightly smaller than the analysis-verified impacts while using a metric based in observation space. This result is in general agreement with the comparison between the OSSENR and OSSEANA case shown here, although notably a similar version of the data assimilation and forecast model was used by Todling. Cardinali (2018) found larger differences between the relative ranking of forecast impacts by data type when comparing an observation-based verification and metric with analysis-verified adjoint estimates of observation impact, with some data types having greater beneficial impacts with observation-based impact, and other data types having smaller beneficial impacts.

While minor differences in the adjoint estimates of observation impact were found between the two verification methods, there could be circumstances under which significant discrepancies are possible. For example, if the adjoint tool were used to study a particular single event, or similarly the use of a small region for calculation of the error norm, there may be greater uncertainty in the adjoint estimates. If the self-analysis is expected to have substantial errors in the particular spatiotemporal region of interest, it may be worthwhile to select instead a different field for verification of the error norm, such as the analysis from a different model, or the use of observational verification instead (Todling 2013; Cardinali 2018). If multiple options for verification are available and the adjoint calculations are not computationally onerous, the adjoint calculations could be repeated with different verification choices to establish the range of uncertainty of the observation impacts.

The ranking of FSOI of various observation types can differ between NWP models and data assimilation systems (Gelaro et al. 2010). These differences are due to the relative weightings of observation types in the data assimilation system and also the characteristics of model error during the initial forecast period. In the OSSE framework, the model error is lower than for real world forecasts using the same model, leading to a reduction in the magnitude of the forecast impacts. For real data with the same forecast model, the uncertainty from using self-analysis verification is therefore likely to be larger than found in this OSSE study. For a different model with different model errors and biases, the discrepancies between verification methods for FSOI may be larger or smaller for various data types and geographical regions.

The results found here indicate that the adjoint estimations of observation impact are robust despite uncertainties in the verification fields used for the calculations of 24-h forecast errors, especially when considering large datasets rather than individual observation impacts. These findings should encourage trust in the use of adjoint models for monitoring and evaluating observation impacts on short-term forecasts.

Acknowledgments. Support for this project was encouraged by Steven Pawson and provided by NASA/GMAO core funding. The software for simulating GPS-RO observations was provided by the Radio Occultation Processing Package (ROPP) of the Radio Occultation Meteorology (ROM) Satellite Applications Facility (SAF) of EUMETSAT, with the assistance of Sean Healy at ECMWF. We thank three anonymous reviewers whose comments led to improvements in the paper.

REFERENCES


