RESEARCH ARTICLE

Addressing the sensitivity of forecast impact to flight path design for targeted observations of extratropical winter storms: A demonstration in an OSSE framework

A. C. Kren[1,2] | L. Cucurull[2] | H. Wang[3,4]

[1]Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL

[2]NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, FL

[3]Cooperative Institute for Research in Environmental Sciences, University of Colorado—Boulder, Boulder, CO

[4]NOAA Earth System Research Laboratories, Boulder, CO

Correspondence
Andrew C. Kren, NOAA Atlantic Oceanographic and Meteorological Laboratory, 4301 Rickenbacker Causeway, Miami, FL, 33149, USA.
Email: andrew.kren@noaa.gov

Funding information
Sensing Hazards with Operational Unmanned Technology, Grant/Award Number: NOAA Weather Satellite Data Mitigation Gap Reserve

Abstract

Few studies have examined the forecast uncertainties brought about from varying aircraft flight track patterns in targeted observations for extratropical winter storms. To examine the degree of uncertainty in downstream forecasts caused by different aircraft flight patterns, a series of observing system simulation experiments (OSSEs) are performed and demonstrated for two extratropical winter storms identified in the European Centre for Medium-Range Weather Forecasts (ECMWF) T511 Nature Run using the National Centers for Environmental Prediction Global Data Assimilation System and Global Forecast System (Q1FY15). Winter storms were chosen to support operational Pacific Ocean targeting strategies using unmanned aircraft. For these two storms, objective and composite flight tracks are generated as they could occur in an operational field mission to sample sensitive areas and meteorologically important regions, and then the changes in downstream forecasts across the various flight tracks are evaluated. The forecast impact downstream is sensitive to flight track orientation and shows case-dependent results, with some flight patterns leading to significant improvements, while others result in neutral to degraded forecasts. The degree of downstream uncertainty in the verification region can vary up to 8% from the different flight paths, depending on the metric used and the atmospheric variables analysed. Although the study is a demonstration of the technique and is limited to only two case studies, it suggests that uncertainty in flight path design should not be neglected in future field missions. Some guidance for mitigating this uncertainty is also discussed.

KEYWORDS

decision support, forecasting systems, numerical weather prediction (NWP), sensitivity analysis, short range, uncertainty analysis

1 | INTRODUCTION

The concept of targeted observing has been used in operational numerical weather prediction (NWP) models since the 1980s across several field campaigns (Szunyogh *et al*., 2000, 2002; Majumdar, 2016). However, the benefits in forecast skill gained from the use of supplementary observations have been small over the past decade, with the most significant improvements found for the prediction of tropical cyclone track (Majumdar, 2016). Improvements in physical parameterizations, higher resolution global and regional NWP models, the use of more sophisticated data-assimilation systems, as well as an increase in the global observing network may have limited the benefits of targeted observations (Langland, 2005; Hamill *et al*., 2013; Majumdar, 2016). It is also possible that data-assimilation systems are generating suboptimal analyses, especially when dealing with precipitating clouds and the potential competing or lack thereof of satellite radiance observations in these areas (McNally, 2002, 2009), thereby limiting the positive benefit from targeted observations.

Recent observing system experiments (OSEs) have evaluated the effectiveness of NASA's unmanned aerial systems (UAS) Global Hawk (GH) during the National Oceanic and Atmospheric Administration's (NOAA) Sensing Hazards with Operational Unmanned Technology (SHOUT) (Black *et al*., 2014; Dunion *et al*., 2018; Wick *et al*., 2018, 2020) campaign (Kren *et al*., 2016, 2018; Christophersen *et al*., 2017, 2018a, 2018b; Sippel *et al*., 2018). These OSEs addressed both winter storms during the NOAA's El Niño rapid response (ENRR) (Dole *et al*., 2018) mission and tropical cyclones during the hurricane rapid response (HRR) campaign.

In addition to the OSEs, one of the goals of the NOAA SHOUT project was to conduct OSSEs to evaluate and test the effectiveness of different UAS sampling strategies to improve the prediction of high-impact weather events. In contrast to the OSEs, which evaluate the impact of existing observations, the OSSEs can be used to investigate the potential impact of current or proposed observing platforms in various observing configurations (Atlas *et al*., 2001; Halliwell *et al*., 2014; Privé *et al*., 2014; Atlas *et al*., 2015; Hoffman and Atlas, 2016; Cucurull *et al*., 2017; Leidner *et al*., 2017).

Flight path design was one of the key components in the SHOUT field campaign. During the real-time SHOUT-ENRR mission (Kren *et al*., 2018), the flight paths created by mission scientists were predominately a combination of both targeting strategies and the meteorological situation, along with other flight restrictions and constraints. This methodology raised a question, namely: What is the uncertainty brought about from the flight path design? While earlier OSSE studies have investigated the sensitivity of

tropical cyclone track to flight path designs (Privé *et al*., 2014; Ryan *et al*., 2015, 2019), to the best of the authors' knowledge, little work has been done to examine the sensitivity inherent in the forecast impact downstream from varying flight-track-pattern deployments. This information would be highly beneficial in future field missions focused on extratropical winter storms.

The objective of the study is to evaluate how changes in the flight path design can affect weather forecasts. It follows an approach similar to that used in field campaigns by combining objective targeting with meteorological reasoning. The study uses predefined verification regions (VRs) across two winter storms over the Pacific Ocean. Furthermore, the long-endurance GH is used to sample a much larger region than in earlier programs (e.g. Szunyogh *et al*., 2000, 2002). Finally, the research is carried out using an OSSE framework so that the impact of multiple flight tracks can be quantitatively evaluated.

The paper is structured as follows. The methodology, including the OSSE, flight track design and experimental set-up, are described in Section 2. Sections 3 and 4 examine the sensitivity of downstream impacts to varying flight patterns for two winter storm cases. Finally, Section 5 discusses the major findings of the study.

2 | METHODOLOGY

A critical component of an OSSE is the nature run (NR), a long-integration weather simulation considered to represent the true atmosphere. In the study, the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS), v.cy31r1 (Masutani *et al*., 2007) is used. The ECMWF's NR has a spectral resolution of T511 (about 40 km) with 91 vertical levels (Masutani *et al*., 2007; Andersson and Masutani, 2010). For more details on the T511 NR, including its representation of climatology, see Masutani *et al*. (2007), Reale *et al*. (2007), Andersson and Masutani (2010), McCarty *et al*. (2012) and Errico *et al*. (2013). The NR has previously been validated and used for other OSSE applications (Masutani *et al*., 2007; Reale *et al*., 2007; Andersson and Masutani, 2010; McCarty *et al*., 2012; Errico *et al*., 2013) and was used to examine the impact of the GH dropsonde data on winter storm forecasts (English *et al*., 2018; Peevey *et al*., 2018).

Perfect conventional and satellite observations are simulated from the ECMWF's T511 NR based on the global observing system of 2012 (Zhu *et al*., 2012). For all the observations assimilated, see Peevey *et al*. (2018, appx A). Although using perfect observations does not mimic operations and is a limitation of the study, the OSSE system was validated by comparing the results with the

dropsonde impact in OSEs from Kren *et al.* (2018) using a similar model configuration. The impact of the GH dropsondes on the operational Global Forecast System (GFS) showed, on average, a reduced error in temperature, relative humidity (RH), wind and sea-level pressure (SLP) by 3–8% relative to the baseline observing system, similar to results shown in the present study. These results were also found by Peevey *et al.* (2018) using the same OSSE system.

A sensitivity map, which can be generated by any targeting method, can provide guidance for designing a flight pattern. In the study, regions of sensitivity for the simulated GH dropsonde observations are generated using the ensemble transform sensitivity (ETS) method (Zhang *et al.*, 2016; Wang *et al.*, 2018). This methodology has been tested in earlier OSSEs (English *et al.*, 2018; Peevey *et al.*, 2018) and implemented and applied in the real-time SHOUT-ENRR field campaign (Kren *et al.*, 2018; Wick *et al.*, 2020). Briefly, the ETS technique is performed using a set of 80 lagged global ensemble forecasts initiated from global ensemble Kalman filter analyses over a 24 hr period, with 20 ensemble members in each 6 hr cycle and centered around five days before a winter storm reaching the VR at the verification time (VT). The error metric uses a dry total energy norm of temperature and zonal and meridional winds at pressure levels of 200, 500 and 700 hPa.

2.1 | Experiment design

Eight OSSE experiments were conducted, covering two winter storms identified in the NR: (1) a winter storm that formed over the central North Pacific and traversed into Oregon by 1200 UTC February 2, 2006 (hereafter OREGON); and (2) a winter storm that formed over the western North Pacific and reached southern Alaska on 0000 UTC February 2, 2006 (hereafter ALASKA). The storms were chosen using four criteria: (1) initial development over the Pacific Ocean, where the SHOUT project had focused for its winter targeting campaign; (2) long-lived, for at least four days; (3) a defined structure with at least one closed contour in the 500 hPa geopotential height and SLP fields; and (4) $\geq$ 4 mm of 6 hr precipitation over the chosen VR.

A control (CTL) experiment for each storm included all simulated perfect observations except the additional dropsondes from the GH. For both storm cases, four simulated flight track experiments were conducted using the perfectly simulated GH dropsondes, interpolated from the NR fields of temperature, specific humidity, and zonal and meridional winds to the specific observation location. The dropsondes contained measurements on 26 fixed levels to include mandatory and significant levels: 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 725, 750, 775, 800, 825, 850, 875, 900, 925, 950, 975 and 1,000 hPa. Vertical levels > 70 hPa were not included due to the approximate GH altitude limit of 19.8 km.

The dropsondes were assimilated similarly to the operational GFS, written to binary universal form for the representation of meteorological data (BUFR), and with no account for dropsonde drift. Accounting for dropsonde drift (Aberson *et al.*, 2017) has recently been implemented in the 2018 operational Hurricane and Weather Research and Forecasting (HWRF) model (Winterbottom, personal communication, 2019), and was not available in the version of the GFS used here.

A summary of the storm cases is provided in Table 1. All simulated tracks take into account the ETS sensitivity and constraints from the GH platform. Therefore, dropsondes can only be released over the open ocean, a total flight time of approximately 24 hr, a maximum of about 80–90 dropsondes in a single mission, and travel is limited to as far west as 175° E due to air traffic control regulations during the SHOUT mission.

To examine the sensitivity of forecasts to operational flight track designs, first an ETS experiment samples only the region of largest sensitivity. The design of the ETS experiments does not include the input from mission scientists or other campaign objectives. In contrast, composite flight patterns are created similar to how they occur in an operational setting, which include taking into account sensitivity signals and the relevant meteorological features and science objectives. The first of these composite flight

TABLE 1   OREGON and ALASKA storm cases

Experiment	Oregon description	Alaska description
CTL	Operational observations only	Operational observations only
ETS	Objective ensemble transform sensitivity (ETS) flight path	Objective ETS flight path
VORT	Sample region of rapidly growing baroclinicity in the vicinity of a low-pressure system	Sample region of rapidly growing baroclinicity in the vicinity of a low-pressure system
JET	Sample jet exit region and ridge over the eastern Pacific	Sample upper-level jet streak over the central North Pacific
MOIST	Sample moisture plume connected to the warm sector	Sample moisture plume connected to the warm sector

patterns samples upper-level jet streaks, denoted as JET. A second flight pattern focuses on atmospheric rivers and moisture, denoted as MOIST. A third experiment targets low-pressure systems and positive vorticity advection, denoted as VORT. Full details on the flight track design are described in Section 2.2. The goal of the paper is not to address the fraction of variability captured by the various flight track designs or whether one flight pattern is more advantageous to sample than another. The objective is to show that sensitivity exists among the different patterns, which could occur in real field missions. The fraction of variability among different paths, however, is something that should be considered in future.

Experiments use the Q1FY15 operational implementation of the National Centers for Environmental Prediction (NCEP) Global Data Assimilation System (GDAS)/ GFS. For details on the Q1FY15 implementation, see McClung (2014). This system uses the three-dimensional variational (3D-Var) configuration with the hybrid ensemble Kalman filter (EnKF) data-assimilation system (Kleist and Ide, 2015a, 2015b). The spectral resolution of the GDAS/GFS (EnKF) system is TL382 (approximately 50 km), while the EnKF component has a resolution of TL254 (about 78 km) with 64 vertical levels. All the forecast runs are initialized 2–3 days in advance of the storm reaching the corresponding VR at the VT in order to mimic the GH targeting capability (48–72 hr before forecast impacts). The VRs consist of $14 \times 14°$ horizontal boxes to cover an envelope of the storm impacts. Throughout the paper, several terms are used when presenting the research results (Table 2).

2.2 | Flight track design and forecast skill metrics

The objective and composite flight tracks were made for the OREGON and ALASKA storms. For the objective flight track design, an algorithm was designed and implemented that accounted for both the GH flight capabilities and ETS sensitivity. Using the average 48–72 hr lead time ETS sensitivity (the average was chosen to account for the change in sensitivity with time during a GH mission), the track software created an initial waypoint (denoted as x_i; bounded by 175° E–125° W in longitude and 19–61° N in latitude) over the Pacific Ocean, in a region of maximum sensitivity. The initial location adheres to the bounds of the approved accessible airspace during SHOUT, as well as the constraint that this location is ≤ 7 hr from Edwards Air Force Base (AFB), the location used during the SHOUT-ENRR. If there were multiple regions with the same maximum ETS, the software selects the first location. The 7 hr time threshold

TABLE 2 Definitions of several terms used throughout the paper

Experiment	Description
Targeting time	Time at which targeted observations are deployed. Units of YYYYMMDDHH
Targeting (sampling) region	Region where targeted observations are deployed
Ensemble transform sensitivity (ETS) sensitivity	Region where error growth may amplify with time, predicted by the ensemble transform sensitivity method
Verification time (date)	Time when a high-impact storm is predicted to reach the verification region. Units of YYYYMMDDHH
Verification region	Region denoting where (1) a high-impact weather system will propagate and (2) anticipated impacts of meteorological significance
Lead time	Hours after a forecast initialization time. Also used to refer to the time before a storm moves into a specified region. Units of hours

was somewhat arbitrary, but was chosen to maximize the flight time over regions prone to larger error growth. It was also an approximate time for the GH, with 335 kn or 172 m·s^{-1} flight-level speed to reach most of the central and eastern Pacific.

Subsequent waypoints (x_n) for the objective flight track were created by calculating the mathematical gradient from the sensitivity map (on a $1 \times 1°$ latitude– longitude grid) at four grid points ($Y = y_1, y_2, y_3, y_4$; square) surrounding the current location (denoted as x_c). The four grid points were separated by $\pm 4°$ from the current waypoint. The gradient ($\Delta; Y - x_c$) was used to find the local maximum in the ETS at the four quadrants. For example, assuming a current (x_c) waypoint ETS of 5 and ETS values (Y) at the surrounding north, south, east and west grid points of 3, 7, 9 and 4, respectively, the algorithm will create a future waypoint to the east, since the gradient of each quadrant is −2, 2, 4 and −1; the local maximum ETS is to the east. If a proposed quadrant is outside the sampling domain and/or over land, these points are excluded due to Federal Aviation Administration (FAA) flight restrictions.

Once a waypoint had been used, it was excluded from future use. In operations, it was normal to repeat a region more than once, but this case was not considered here. When the algorithm reached the specified time limit over the sampling region (here chosen to be 17 hr), it created a final waypoint at Edwards AFB to terminate the flight

track. Dropsonde locations were then created at a density of ±1° latitude–longitude starting with the initial waypoint in the primary proposed sampling region. Less dense dropsonde locations were also made to/from Edwards AFB between the initial and final waypoints. The limitation of this automated algorithm will be discussed below.

In addition to the objectively determined flight tracks, composite flight paths were manually created as they would occur in actual field campaigns to account for not only the sensitivity but also the important meteorological features, such as JET, regions of rapid cyclogenesis, frontal boundaries and regions near atmospheric rivers (Ralph *et al.*, 2005). For example, during SHOUT-ENRR (Kren *et al.*, 2018), one of the targeted winter storms sampled on February 21–22, 2016, included an intensifying extratropical cyclone with an adjacent atmospheric river over the central North Pacific that moved into the Gulf of Alaska (Dunion *et al.*, 2018). In addition to relying on sensitivity analyses, mission scientists also co-ordinated with other aircraft platforms, including the Air Force C-130 J and the NOAA G-IV, both sampling different synoptic regions.

Anomaly correlation (AC) and root mean square error (RMSE) were used as verification metrics. Forecast errors were also examined by using a dry total energy error metric, as described in Peevey *et al.* (2018). The dry total energy error $(\mathrm{m \cdot s^{-1}})$ was calculated as follows:

$$E = \left(\frac{1}{2} \int A \begin{bmatrix} \frac{1}{3}(u_{200}^2 + v_{200}^2 + \frac{c_p}{T_r} t_{200}^2 + \\ \frac{1}{3}(u_{500}^2 + v_{500}^2 + \frac{c_p}{T_r} t_{500}^2 + \\ \frac{1}{3}(u_{700}^2 + v_{700}^2 + \frac{c_p}{T_r} t_{700}^2 \end{bmatrix} \right)^{1/2} \quad (1)$$

where *t*, *u* and *v* represent the differences of analyses and forecasts for temperature and zonal and meridional wind speed over pressure levels of 200, 500 and 700 hPa. This energy error equation is the same as that employed in generating the ETS sensitivity. The terms T_r, A and c_p refer to the reference temperature (300 K), area or domain $(\mathrm{m}^2, \text{VR})$, respectively, over which averaging is performed, and the specific heat of dry air at constant pressure $(1{,}004 \ \mathrm{J \cdot kg^{-1} \cdot K^{-1}})$.
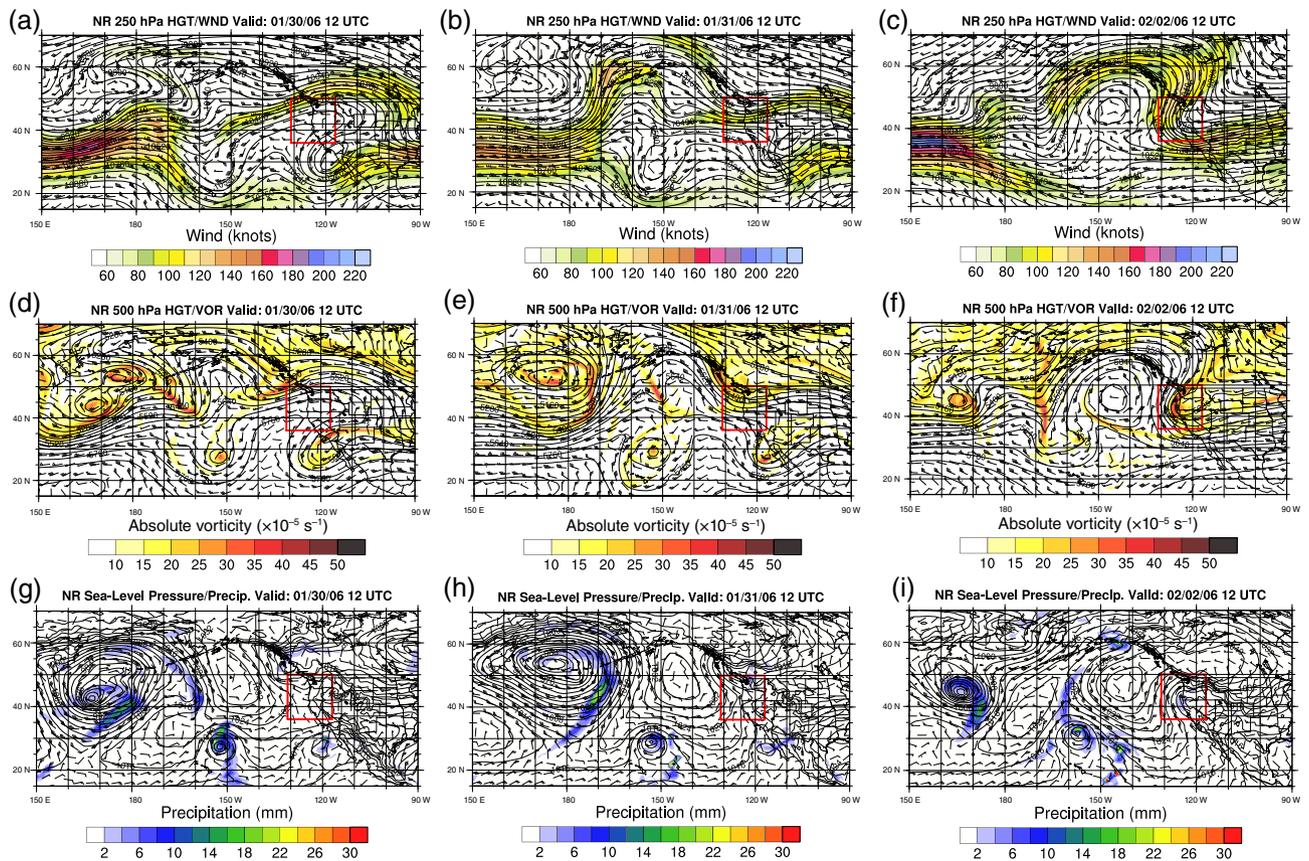
3 | DEMONSTRATION 1: OREGON STORM

3.1 | Synoptic situation

The NR fields of 250 hPa heights and isotachs, 500 hPa heights and absolute vorticity, as well as sea-level

pressure (SLP) and 6 hr precipitation are shown in Figure 1. These fields are plotted for the 72 hr (1200 UTC January 30, 2006) and 48 hr (1200 UTC January 31, 2006) lead times, and at the VT (reaching the domain of Oregon; 1200 UTC February 2, 2006). Data assimilation cycling for OREGON started at 1200 UTC January 30 and ended at 1200 UTC January 31, with forecasts every 6 hr. At the 250 hPa level, a jet streak extending from the eastern North Pacific into Canada at 1200 UTC January 30 (Figure 1a) makes its way southward into the northwest part of the United States at the VT (Figure 1c), with a positively tilted trough in the VR of Oregon at mid-levels (Figure 1f). An initial short wave trough at 500 hPa was located offshore of Washington state and western Canada at 1200 UTC January 30 (Figure 1d). The short wave later interacted with the upper-level jet over western Canada to produce a dip in the jet stream and an area of low pressure at the surface (Figure 1i), with 6 hr precipitation between 2 and 5 mm in the VR.

Figure 2a–i shows the CTL RMSE analyses relative to the NR over the Pacific Ocean and western North America between the 72 hr lead time of 1200 UTC January 30 and the 48 hr lead time of 1200 UTC January 31 for the geopotential height, temperature and wind at the 200, 500 and 700 hPa levels. Three regions exhibit the largest analysis error. First, a region from the western North Pacific into the central and north-central Pacific (Figure 2a–i), in connection with the JET and mid-level trough pattern (Figure 1a–f). Second, a region offshore of the Pacific Northwest and in the verification domain of Oregon, tied to the developing low-pressure system and short wave. Finally, the increased error north and east of Hawaii in connection to a cut-off low and tropical convection (Figure 1g–i).

Figure 3 shows the average 48–72 hr lead-time ETS sensitivity generated approximately five days in advance of the storm reaching the Oregon VR at the VT, along with the corresponding objective ETS simulated GH flight track, as well as the three composite flight tracks of VORT, JET and MOIST. For a targeting time centered on 0000 UTC January 31, highest sensitivity regions over ocean are located upstream over the central North Pacific, south of the Aleutian Islands and over the Bering Sea (Figure 3a). These regions are likely tied to the storm system south of the Bering Sea (Figure 1d), in concert with the exit region of the upper-level jet in Figure 1a, as well as the frontal boundary (Figure 1g) draped from northwest to southeast from the Aleutians into the central North Pacific. These areas are also tied to the analysis errors in the CTL experiment shown in Figure 2, primarily in the temperature and wind fields. The average ETS does not

FIGURE 1 (a–c) Nature run (NR) fields of 250 hPa heights (m) and isotachs (kn) at the (a) 72 hr lead time (1200 UTC January 30, 2006), (b) 48 hr lead time (1200 UTC January 31, 2006) and (c) VT (1200 UTC February 2, 2006) for the OREGON observing system simulation experiment (OSSE) storm case. (d–f) Same results as for (a–c), but for 500 hPa heights and absolute vorticity ($\times10^{-5}$ s^{-1}). (g–i) Same results as for (a–c), but for sea-level pressure (SLP) and 6 hr precipitation. Red boxes denote the verification region over Oregon state, covering a region from 36–50° N to 131–117° W. contour intervals are 60 m, 10 kn, 5 s^{-1} and 4 hPa for heights, isotachs, absolute vorticity and SLP, respectively

capture the large analysis errors over the western North Pacific in connection to the jet and trough pattern. This is discussed in more detail in Section 4.1 with the ALASKA case.

For comparison with the ETS sensitivity, Figure 4 shows the mean ensemble standard deviation computed from the ensemble forecasts valid at the targeting time of 0000 UTC January 31 for height, temperature and wind at the 200, 500 and 700 hPa levels. The ETS sensitivity captures most of the ensemble spread at the targeting time over the Russian continent, the Bering Sea and the central North Pacific.

Based on the largest sensitivity region present over the central North Pacific, the ETS flight track sampled along the path of the frontal boundary and exit region of the upper-level jet in connection to the Bering Sea storm system. Although this storm does not directly impact Oregon, it is possible that the error sensitivity that exists in this region influences the downstream pattern, track

and intensity of the storm that directly influences the state of Oregon. The large ETS sensitivity over Russia was not sampled as it was outside the approved accessible airspace.

In addition to the objective ETS experiment, Figure 3b–d shows three composite flight tracks. A VORT composite flight experiment (Figure 3b) is created to sample the positive vorticity region and baroclinic development in connection with the storm system that traversed southward into Oregon at the VT, as well as the ETS sensitivity around 140° W and 45° N. This region also encompasses the analysis error and ensemble spread seen in the CTL (Figures 2 and 4).

A JET composite flight experiment (Figure 3c) depicts a butterfly pattern that is designed to sample the exit region of the upper-level jet stream east of 180° E, a region of sensitivity in the Bering Sea and Gulf of Alaska, and the ridge and jet off the coast of the United States. Finally, a MOIST composite flight focuses on the region of moisture
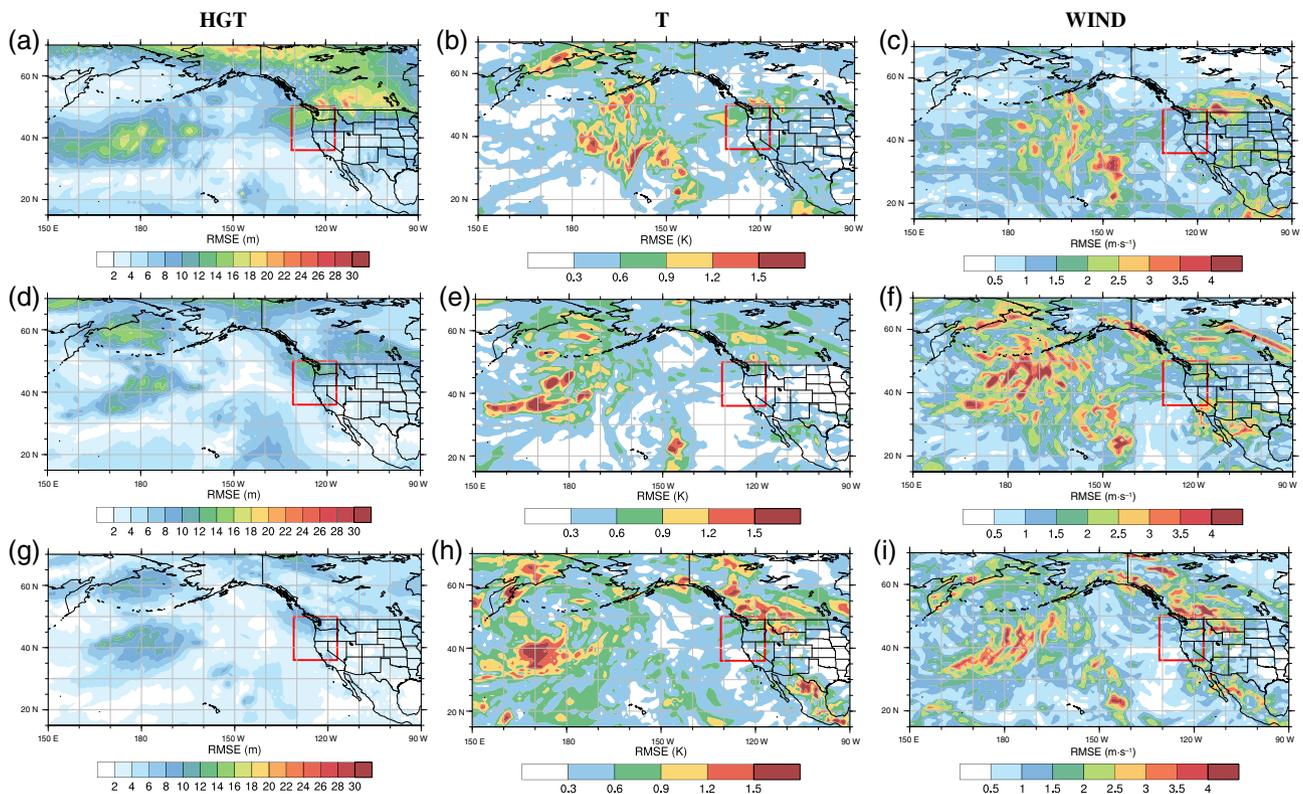
FIGURE 2 Analysis root mean square error (RMSE) in the control (CTL) experiment, verified against the nature run (NR), for the OREGON case between 1200 UTC January 30 and 1200 UTC January 31, 2006. Results are shown for geopotential height (a, d, g), temperature (b, e, h) and wind (c, f, i) for the (a–c) 200, (d–f) 500 and (g–i) 700 hPa levels. Red boxes denote the verification region (Figure 1). Contour intervals are 2 m, 0.3 K and 0.5 m·s^{-1} for geopotential heights, temperature and wind, respectively

and atmospheric river over the central North Pacific and the sensitivity south of the Aleutians (Figure 3d).

3.2 | Results

3.2.1 | Analysis error relative to control forecasts

Before showing forecast impacts from the individual flight experiments, Figure 5 shows the analysis error differences between each flight path and CTL in terms of the dry total energy error (Equation 1). Results are averaged over the five cycles of dropsonde assimilation from 1200 UTC January 30 to 1200 UTC January 31. All dropsonde experiments reduce the analysis error relative to the CTL, with the largest reduction found in the ETS experiment and the smallest reduction in the VORT experiment. The reduced analysis errors are largely in the vicinity of where the dropsondes are assimilated (Figure 3). The subsequent sections will show the uncertainty in downstream forecast impacts for the OREGON

case, that can arise from varying flight track pattern deployments.

3.2.2 | Anomaly correlation and root mean square error

Figure 6a, b shows the AC difference (experiments minus the CTL) for the four flight paths shown in Figure 3. Forecasts are averaged over all assimilation cycles and presented over the VR of Oregon for 500 hPa geopotential heights and SLP as a function of forecast lead time. Overall, while most flight paths show higher forecast skill in SLP (Figure 6a) and 500 hPa geopotential heights (Figure 6b), the improvement over CTL is not statistically significant. However, some sensitivity to the flight tracks are still evident from the figures. JET and MOIST show a statistically significant improvement over CTL after 72 hr in SLP, and JET shows a significant improvement over CTL for the 500 hPa geopotential heights. The other flight tracks show neutral impact relative to the CTL, stressing the variability of the results depending on flight pattern.
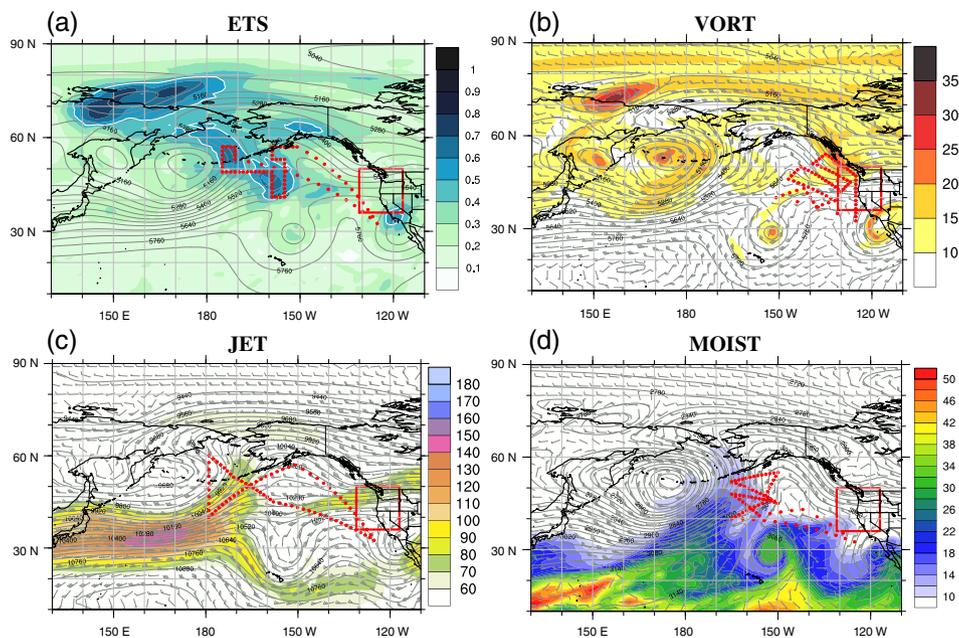
FIGURE 3 Simulated flight tracks for the OREGON observing system simulation experiment (OSSE) storm, as in Table 1, from four cycles between 1800 UTC January 28 and 1200 UTC January 29, 2006. (a) Ensemble transform sensitivity (ETS): normalized ETS sensitivity (shaded contours; values of 0.5 are contoured in white) and 500 hPa geopotential height (grey contours; m). (b) Absolute vorticity (VORT): 500 hPa heights and absolute vorticity ($\times 10^{-5}$ s^{-1}). (c) Upper-level jet streaks (JET): 250 hPa heights and isotachs (kn; shaded). (d) Atmospheric rivers and moisture (MOIST): 700 hPa height, wind and integrated water vapor (IWV; mm, shaded). All fields are averaged over the 48–72 hr lead time, showing a targeting time of 0000 UTC January 31, 2006, and verification time (VT) of 1200 UTC February 2, 2006. Red dots denote the dropsonde release locations of the Global Hawk (GH). Red boxes denote the verification region as in Figure 1

Figure 7a, b shows the same results but averaged over the contiguous United States (CONUS; 25–50° N, 65–125° W). As with the VR, there is a statistically significant improvement over CTL after 72 hr for both the JET and MOIST experiments for SLP (Figure 7a) as well as in the ETS (between 60 and 84 hr) and MOIST (first 60 hr of lead time) experiments for the 500 hPa heights (Figure 7b). The VORT flight pattern experiment does not improve results over the CONUS.

In addition to AC skill, the *per* cent improvement in the RMSE relative to the CTL, as in Campins *et al.* (2013), is examined. Figure 8a, b shows the relative RMSE (positive denotes improvement from dropsondes) averaged over all assimilation cycles and over the Oregon VR for SLP and the 500 hPa geopotential heights for all flight experiments. Overall, neutral results are found for SLP RMSE (Figure 8a), with the exception of a statistically significant degradation in the first 18 hr for the VORT experiment and an improvement from 36 to 60 hr in the ETS path. For 500 hPa geopotential heights (Figure 8b), all experiments provide a statistically significant improvement over CTL in the first 12 hr. This improvement ranges between 1% and 4%. The ETS path provides an additional improvement of 2% over CTL between 36 and 84 hr, consistent with the AC in Figure 7b. These

results indicate that, depending on the atmospheric variables and metric chosen, some flight paths lead to a significant improvement, whereas others are neutral, consistent with the AC results in Figures 6 and 7 when looking at the downstream uncertainty from the flight path design.

Figure 9a–c shows the relative RMSE as a percentage for the different experiments relative to the CTL for geopotential height, zonal and meridional winds, RH, and temperature at vertical levels of 200–925 hPa. The results are averaged over all forecast cycles and forecast lead times of 0–96 hr over the VR. Generally, the addition of added dropsondes improves the downstream forecast (neutral to 2%) for most variables and pressure levels. The impact is greatest in the upper-level fields, similar to Irvine *et al.* (2009). The ETS and VORT experiments consistently reduce the RMSE relative to the CTL for all variables, while the MOIST experiment shows degradation in upper-level geopotential heights and low-level temperature. The JET path also shows some degradation in 200 hPa RH. The results found for this case show that varying the flight pattern can impact whether the downstream is improved or degraded over both a smaller VR or a larger region such as the CONUS.
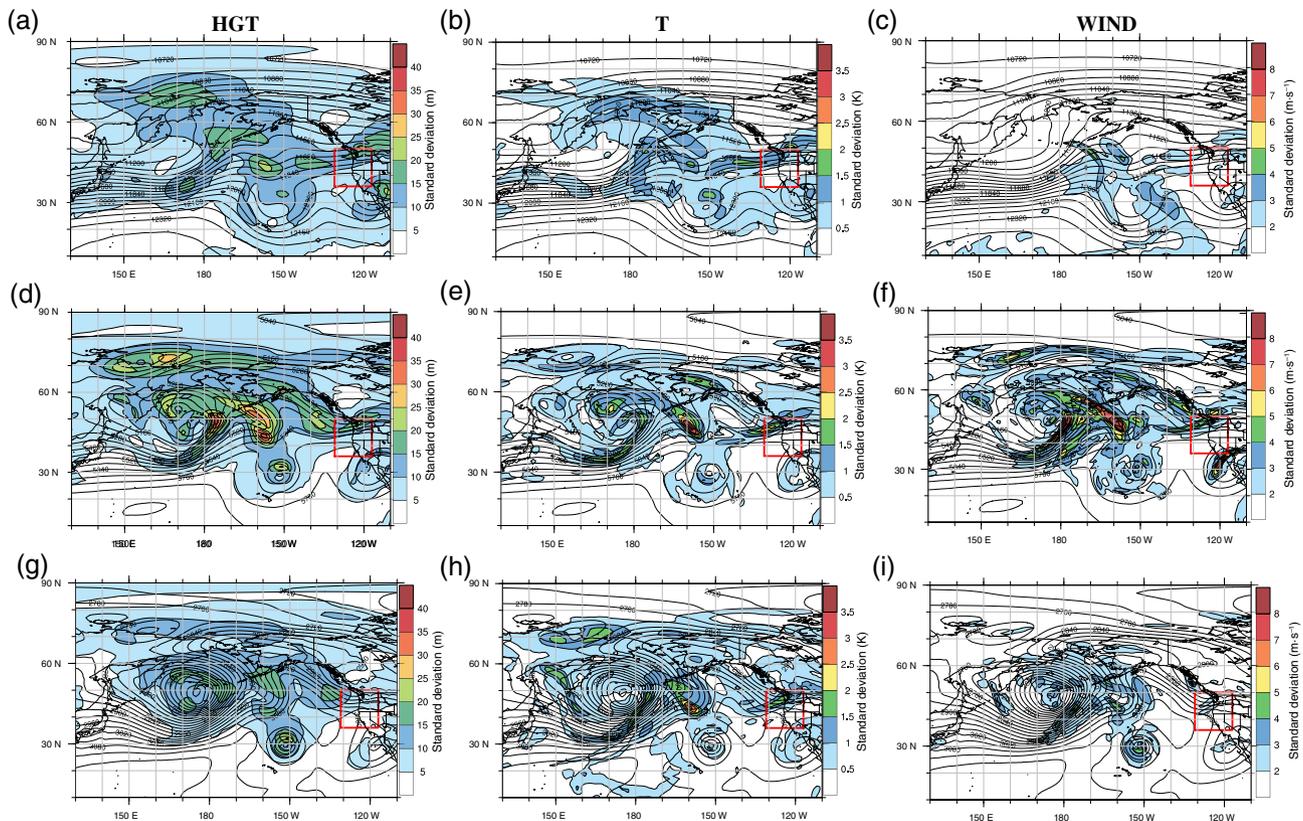
FIGURE 4 As for Figure 2, but for the mean ensemble standard deviation (shaded) and geopotential heights (contour lines) across 80 ensemble members from four cycles between 1800 UTC January 28 and 1200 UTC January 29, 2006. Results are valid at the targeting time of 0000 UTC January 31, 2006. Contour interval for geopotential heights are every 80, 60 and 30 m for the 200, 500 and 700 hPa levels, respectively

3.2.3 | Energy error and propagation of targeted signal into the VR

The previous section showed how varying flight patterns for the OREGON case can lead to uncertainty in downstream weather. In this section, results are further verified in terms of a dry total energy error (Equation 1) in order to provide a more comprehensive evaluation.

Figure 10a shows the *per* cent change in total energy error for each dropsonde experiment, relative to the CTL, as a function of verification date over the VR, along with 95% confidence intervals. Here, a negative change implies less error in the dropsonde experiments compared with the CTL. Consistent with Figures 6–8, altering the flight path design changes the downstream weather pattern. That is most evident in the ETS experiment: a statistically significant improvement (1%) over CTL is found from 1200 UTC February 1 to 0000 UTC 3 February (Figure 10a), whereas the other three experiments are largely neutral. Many of the lead times for the dropsonde experiments, except for the ETS path, fall along the 1–1 line (Figure 10b).

An example is additionally provided to show how sampling two slightly different targeting regions influences Oregon and other downstream areas and can lead to uncertainty in the weather forecast. Energy error differences between the analyses and forecasts of the objective ETS and CTL experiments are computed relative to the NR. The relative improvement of the ETS experiment is compared with the CTL using the difference field at each time step (ETS minus the CTL). An improvement in the objective ETS experiment over CTL is represented by negative differences (reduced error).

Figure 11 shows the energy error differences for a single GFS forecast initialized at the 2.5 day lead time of 0000 UTC January 31, midway through the GH flight track. At the analysis time (Figure 11a), negative energy error differences are present along the frontal boundary, from the southern Aleutians, southeast into the central North Pacific just north of 40° N, indicating that simulated GH dropsondes improve the analysis. This reduced energy error signal propagates northeastward into the eastern Bering Sea by 1200 UTC January 31 (Figure 11b), and into northern Alaska and the Gulf of Alaska by 0000
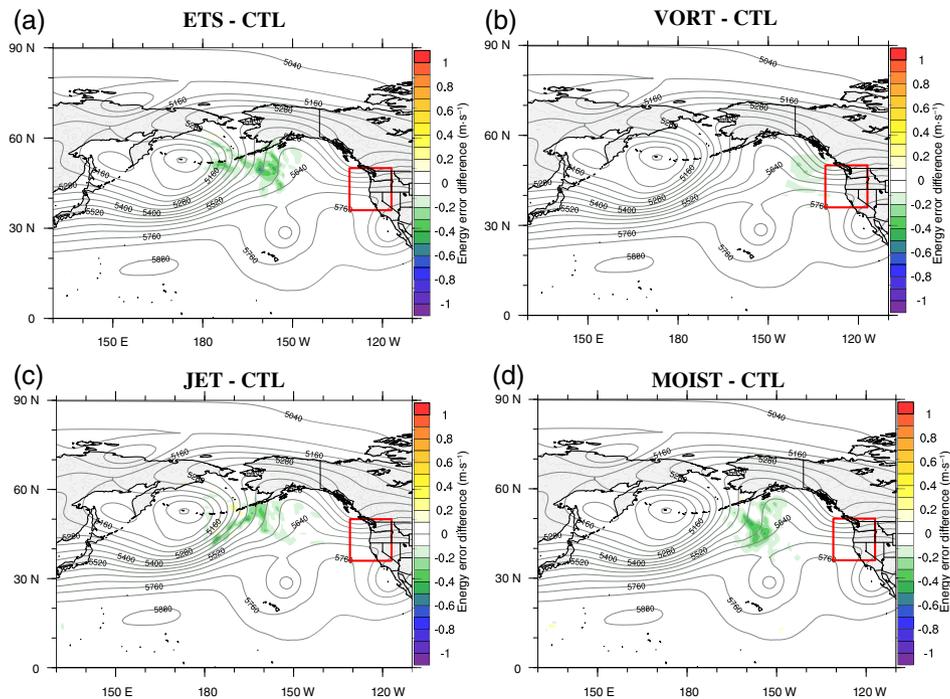
FIGURE 5 Analysis error difference (using total energy error, shaded; m·s⁻¹) relative to the control (CTL) experiment (flight experiment minus the CTL) and with respect to the nature run, averaged over the five cycles of dropsonde assimilation (1200 UTC January 30–1200 UTC January 31, 2006) for the OREGON storm. Results are shown for: (a) ensemble transform sensitivity (ETS) minus the CTL; (b) Absolute vorticity (VORT) minus the CTL; (c) upper-level jet streaks (JET) minus the CTL; and (d) atmospheric rivers and moisture (MOIST) minus the CTL. Negative differences indicate a reduction in analysis error relative to the CTL. Overlaid are the nature run mean 500 hPa geopotential heights over the five cycles. The verification region is indicated as shown in Figure 1

UTC February 1 (Figure 11c). A secondary signal forms south of the boundary and north of the cut-off low-pressure system at 30° N and 155° W by 0000 UTC February 1 (Figure 11c). This secondary signal later results in large downstream positive impacts over the VR, southern California and western Mexico (Figure 11f).

At 1200 UTC February 1 (Figure 11d), the energy error differences are present in three primary locations: (1) over the Bering Sea in association with the mid-level trough; (2) over the central Gulf of Alaska tied to the mid-level ridge axis; and (3) south of the ridge between 20 and 40° N in the east-central Pacific in connection with the cut-off mid-level low-pressure system. While overall these areas are indicative of reduced energy error, some degradation is present along 145° W and 40° N, as well as between 150 and 125° W in a narrow north–south region between 20 and 30° N. Between 0000 and 1200 UTC February 2 (Figure 11e, f), the signal over the Gulf of Alaska is advected southeastward into the VR of Oregon state, resulting in a reduction in forecast error.

It is unclear whether the signal relevant to the VR appears to wash out by 1200 UTC February 1, since the signal present near the Gulf of Alaska ridge dissipates

by 0000 and 1200 UTC February 2. The primary signal that advects into Oregon may be a combination of the cut-off low region and the high pressure south of 45° N and east of 155° W. In this example, most of the sampled region along the frontal boundary penetrates up into Alaska with the mid-level flow being southwesterly. The blocking high over the eastern North Pacific prevents much of the signal from reaching Oregon, consistent with prior work showing the difficulty of targeting methods to account for blocked flows (Sellwood et al., 2008; Majumdar et al., 2010). One option would be additionally to sample areas downstream of the ridge. Finally, a southern branch of the sampled region becomes entrenched into the cut-off low, allowing downstream effects to reach southern California and Mexico.

As a comparison with Figure 11, Figure 12 shows the same results, but comparing the outcome from two different flight patterns, JET minus ETS. At the analysis time (Figure 12a), a reduced energy error relative to the ETS experiment is closer to the jet region between 40 and 50° N and between 180° E and 170° W, while a higher energy error is present along the frontal boundary further north near Alaska. These differences can be explained by the

FIGURE 6 Anomaly correlation difference (flight experiment minus the control (CTL) experiment; solid lines) averaged over all assimilation cycles and over both the (a, b) Oregon and (c, d) Alaskan verification regions (VRs) for the OREGON and ALASKA storms for the ensemble transform sensitivity (ETS) (red), absolute vorticity (VORT) (green), upper-level jet streaks (JET) (orange) and atmospheric rivers and moisture (MOIST) (blue) cases relative to the nature run (NR) for (a, c) sea-level pressure (SLP) and (b, d) 500 hPa geopotential heights. Overlaid are the 95% confidence intervals (dashed lines) using a paired t-test. Differences are significant at the 95% level when the mean difference exceeds the confidence interval. Results are shown as a function of forecast lead time. The vertical black lines on each plot denote the 24 and 72 hr forecast lead times, the primary lead times focused on during the SHOUT campaign
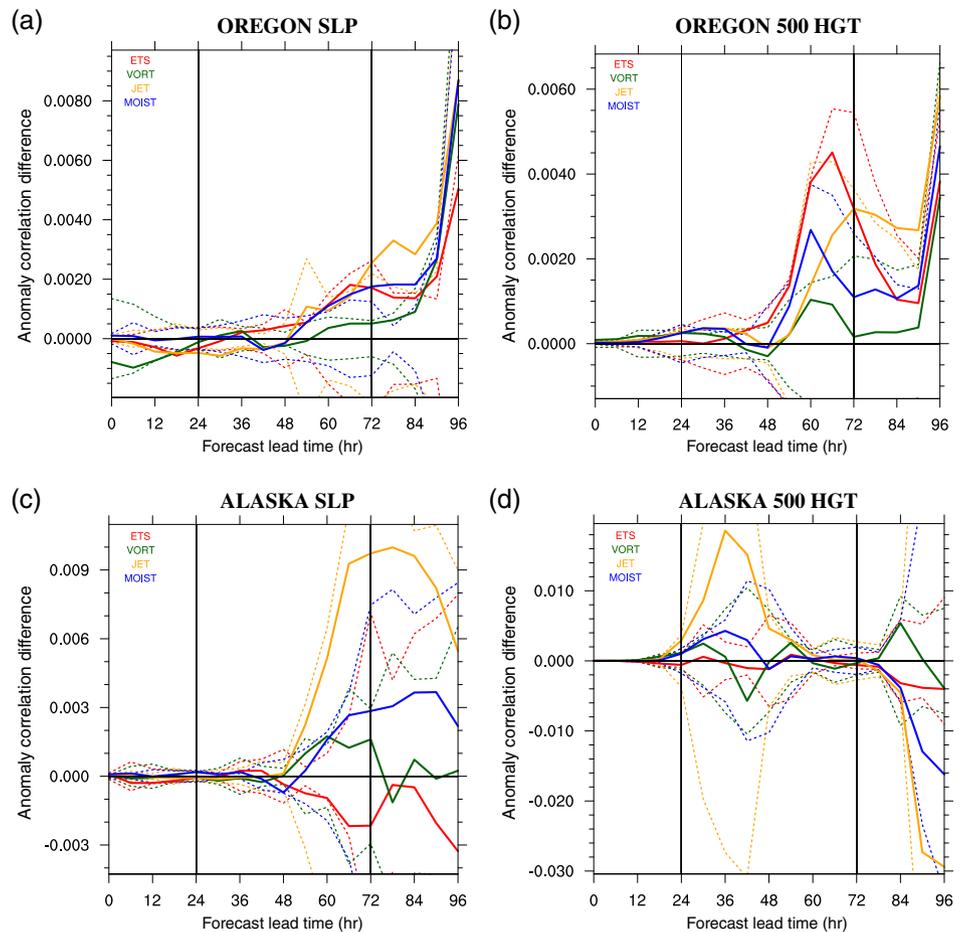
varying dropsonde release locations in this data assimilation window. As the forecast progresses, the pattern of downstream impacts differs, with the JET experiment indicating more impact over the eastern North Pacific in the vicinity of the ridge (Figure 12c–e), and largely an increase in forecast error relative to ETS at the VR by 1200 UTC February 2 (Figure 12f). This comparison shows how slight deviations in the sampling region, brought about by varying flight path design, lead to downstream uncertainty.

Table 3 shows the number of model forecasts at the chosen VR over Oregon, at the VT of 1200 UTC February 2, with reduced error relative to the CTL and relative to the total number of forecasts. The results are stratified by energy error and an RMSE of 500 hPa geopotential height and SLP. Consistent with Figure 10, the ETS experiment improves model forecasts relative to the CTL 100% of the time for the three metrics. However, deviations in the flight path, as indicated by the VORT, JET and MOIST experiments, indicate a range of 20–80% improvement over CTL depending on the metric chosen. This further emphasizes the sensitivity that comes from operational flight planning. Slight changes in the targeting area will alter the downstream results.

3.2.4 | Precipitation verification

Figure 13a shows the NR 72 hr-accumulated precipitation from the start of model cycling of 1200 UTC January 30 to the VT of 1200 UTC February 2. The OREGON case produced a total precipitation of 20–25 mm over western Oregon state. Figure 13c shows the corresponding equitable threat score (EQTS) (Wilks, 2006) computed over the Oregon verification domain for various thresholds of 24 hr-accumulated precipitation (from 0.2 to 15 mm) and for forecast times of 24 to 48 hr. The EQTS is comparable between CTL and the four flight experiments, indicating no significant differences in precipitation forecasts for this particular case.

4 | DEMONSTRATION 2: ALASKA STORM

4.1 | Synoptic situation

Figure 14 shows the NR fields of 250 hPa heights and isotachs, 500 hPa heights and absolute vorticity, and SLP and 6 hr precipitation for the 72 hr (0000 UTC January
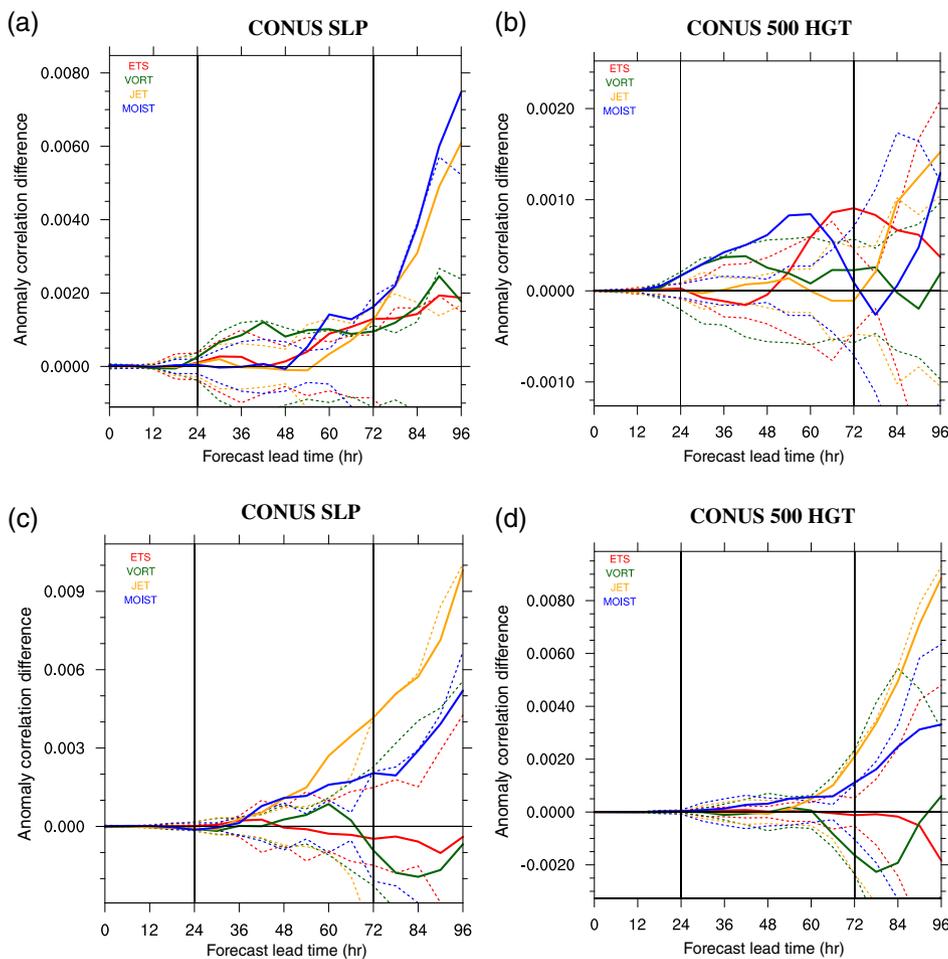
FIGURE 7   As for Figure 6, but for the contiguous United States (CONUS; 25–50° N, 65–125° W) region

30, 2006) and 48 hr (0000 UTC January 31, 2006) lead times, and at the VT (reaching the domain of Alaska; 0000 UTC February 2, 2006). Data assimilation cycling for the ALASKA storm started at 0000 UTC January 30 and ended at 0000 UTC January 31. The system that influenced Alaska was a short wave over the western North Pacific around 40° N (Figure 14d) at 0000 UTC January 30, connected to a strong JET of 160 kn (Figure 14a). The short wave later interacted with the downstream low-pressure system near the western Aleutians (Figure 14d, e) and tracked northeastward into the Bering Sea, impacting the VR of Alaska with up to 18 mm of 6 hr precipitation (0000 UTC February 2) (Figure 14i).

Figure 15a–i shows the CTL RMSE analysis over the Pacific Ocean and western North America from the 72 hr lead time of 0000 UTC January 30 to the 48 hr lead time of 0000 UTC January 31 for geopotential height, temperature and wind at the 200, 500 and 700 hPa levels. Most analysis errors can be traced to the ALASKA storm (Figure 14) along the western North Pacific into the Bering Sea. The errors maximize along the storm track, with the largest errors occurring in the 500 and 700 hPa

levels of up to 1.5 K in temperature, 18 m in height and 4 m·s^{-1} in wind speed (Figure 15d–i).

Figure 16a shows the average 2–3 day lead time ETS sensitivity generated approximately five days in advance of the storm reaching Alaska at the VT, along with the ETS-simulated GH flight track, and VORT, JET and MOIST experiments. The sensitivity study was centered on a targeting time of 1200 UTC January 30.

The largest signal shown in the average ETS map in the upstream Alaska area is over a region from 170 to 150° W and from 35 to 55° N. This region is tied to the JET presented in Figure 14b, as this is collocated with the left exit region of the jet where dynamic instability can lead to forecast uncertainty. The sensitivity is also tied to the frontal boundary draped northwest to southeast south of the Aleutians and the analysis errors present in Figure 15. An additional location of sensitivity exists in the vicinity of the two low-pressure systems in the western North Pacific (Figure 16a), which is also seen in the mean ensemble spread in Figure 17 and analysis errors in Figure 15. Further examination of the ETS maps at each 6 hr time period found that there are noticeable sensitivity signals related to the two low-pressure systems

FIGURE 8 Relative root mean square error (RMSE) (solid lines are expressed as percentages, where positive values indicate improvement relative to the control (CTL) experiment) averaged over all assimilation cycles and over both the (a, b) Oregon verification regions (VRs) and (c, d) Alaska VR for the OREGON and ALASKA storms for all case study experiments relative to the nature run (NR) for (a, c) sea-level pressure (SLP) and (b, d) 500 hPa geopotential heights; results shown as a function of forecast lead time. Colors and the legend denote the appropriate experiments in each storm case. Overlaid are the 95% confidence intervals (dashed lines), as in Figure 6
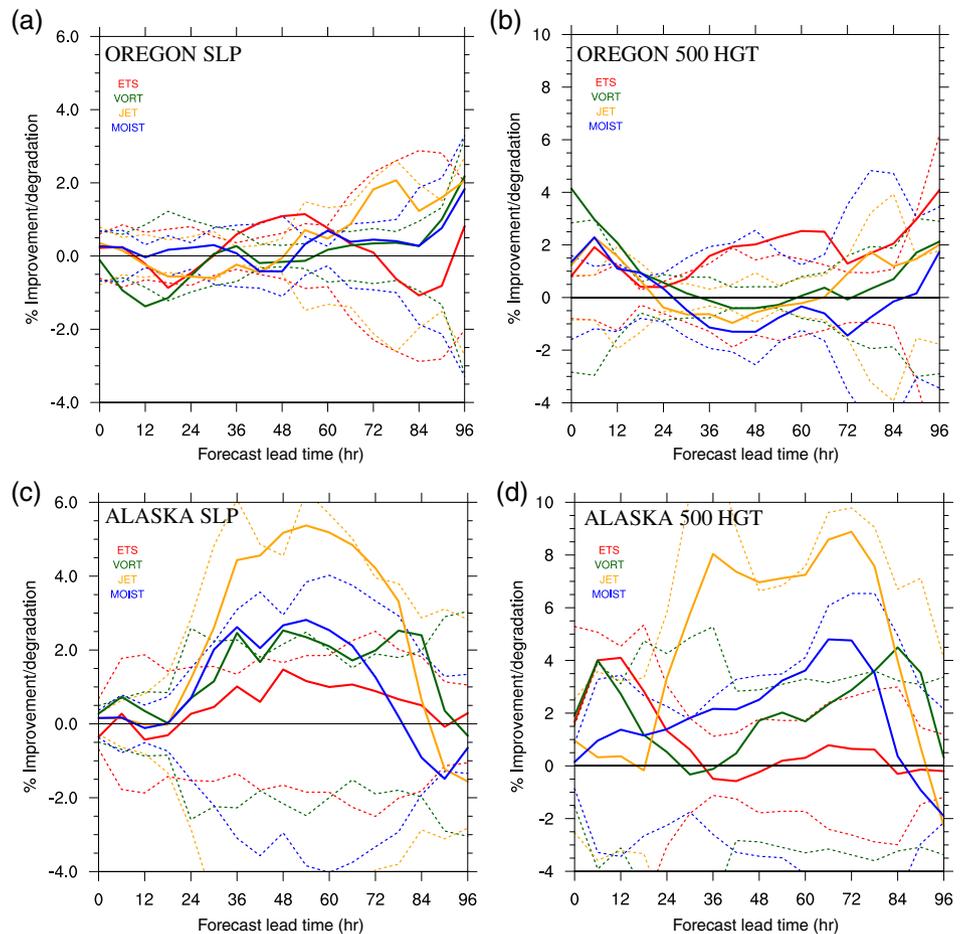
and the jet stream. However, this is not reflected in the ETS signal in Figure 16a for the flight design configuration because the ETS signal was averaged over the five assimilation time windows. This averaging filtered out some signals from the faster propagating flow of the low-pressure system and jet stream, and missing some of the important features over the western North Pacific. Another important reason for the average ETS not fully capturing the analysis errors in Figure 15 is related to the ensemble spread (Figure 17). The analysis errors emphasize an extended area upstream of Alaska, whereas the ensemble spread puts emphasis on the trough that is in close proximity to Alaska.

Figure 16b–d shows the three composite flight tracks based on the synoptic pattern and the sensitivity analysis. The VORT experiment (Figure 16b) is designed to sample the secondary region of ETS sensitivity depicted in Figure 16a that is tied to the two low-pressure systems (Figure 14e), as well as the frontal boundary east of 180° E and south of 50° N. This flight track is also in a region of large temperature and wind RMSE as shown in Figure 15.

The JET composite flight track (Figure 16c) is similar to the OREGON case (Figure 3c), in which a butterfly

pattern is created to sample the exit region of the jet streak from west to east along 170° E to 160° W between 30 and 45° N, as well as the sensitivity region in the central North Pacific. Although the region sampled with this flight track did not exhibit much sensitivity, this region was chosen due to dynamical instability and the presence of the larger errors and ensemble spread seen in Figures 15 and 17. A final experiment, called MOIST (Figure 16d), samples the atmospheric river between 180 and 150° W.

4.2 | Results

4.2.1 | Analysis error relative to control forecasts

As with OREGON, Figure 18 shows the analysis error differences between each flight path and CTL in terms of the dry total energy error during the dropsonde assimilation period (0000 UTC January 30–0000 UTC January 31). All dropsonde experiments reduce the analysis error relative to the CTL, with the largest reduction found in the JET experiment in the vicinity of the JET where the dropsondes were assimilated (Figure 16c).
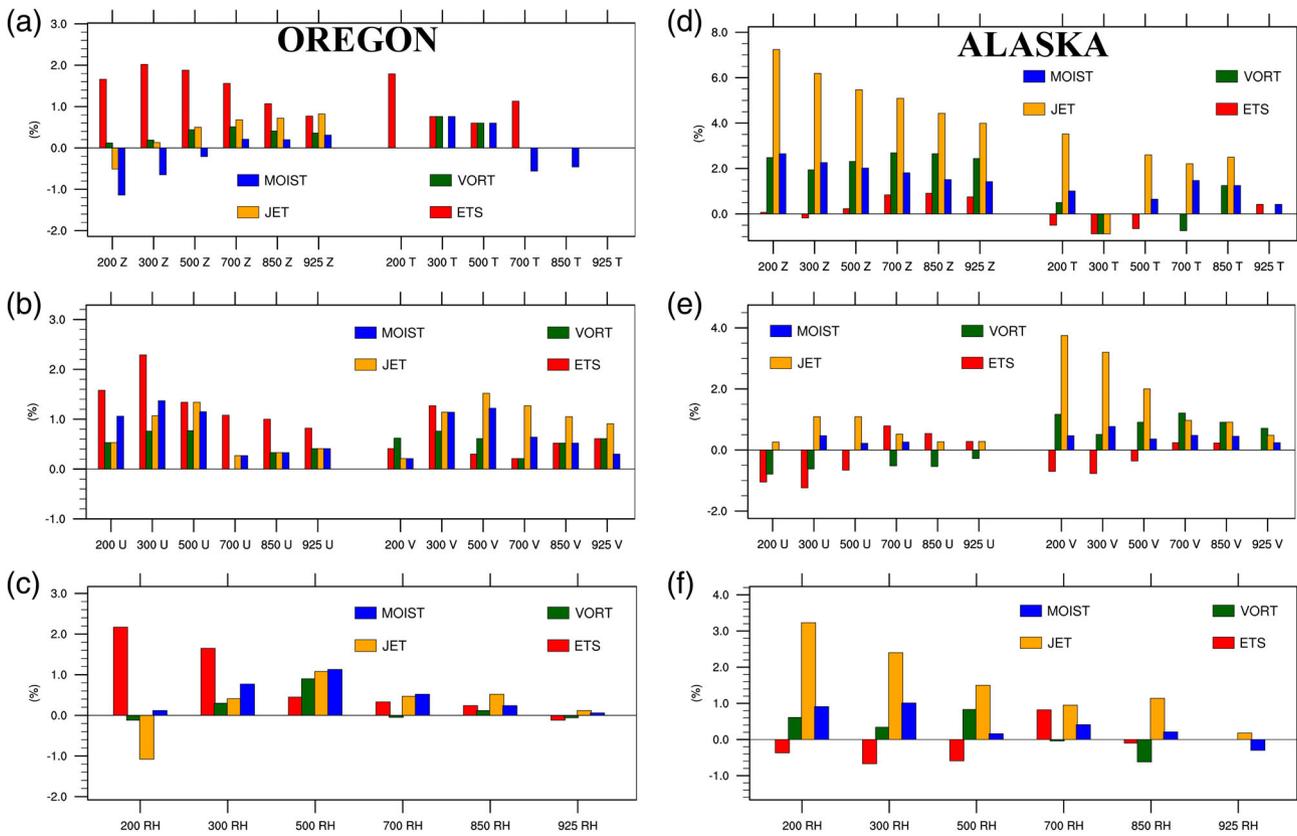
FIGURE 9 Relative root mean square error (RMSE) (%) of all flight track experiments for the (a–c) OREGON and (d–f) ALASKA cases for variables of geopotential height (*Z*; a, d), temperature (*T*; a, d), zonal and meridional winds (*U* and *V*; b, e), and relative humidity (RH; c, f), averaged over all forecast cycles and all forecast lead times up to 96 hr with respect to the control (CTL) experiment, and verified against the nature run (NR). Verification for all variables is shown at levels 200, 300, 500, 700, 850 and 925 hPa. Positive values indicate the reduction in error of each experiment compared with the CTL. The legend denotes the appropriate flight experiments for each observing system simulation experiment (OSSE) storm case. Results are averaged over the respective verification regions (VRs) of Oregon and Alaska

## 4.2.2 | Anomaly correlation and root mean square error

As in Section 3.2.3, the AC difference of each experiment minus the CTL is shown when averaged over all assimilation cycles. In both SLP and 500 hPa geopotential heights, the JET experiment shows the largest improvement over CTL forecasts over the 24–72 hr range (Figure 6c, d). The VORT and MOIST also show improvement, with the ETS flight being largely neutral. However, differences are not statistically significant. Similar to the OREGON case, Figure 7c, d shows the AC differences over the CONUS. In this case, statistically significant improvements are found downstream with the JET and MOIST experiments between 36 and 72 hr in the SLP field. Neutral to degraded forecast skill is found in the VORT and ETS dropsonde experiments. Thus, in terms of AC, the different flight paths lead to forecast uncertainty over the CONUS domain and not over the VR.

Figure 8c, d shows the relative RMSE of each experiment compared with the CTL over the VR. While the

JET experiment shows the largest improvement over CTL of 5–8% in the 24–72 hr time frame for SLP and 500 hPa geopotential heights, the reduced RSME is largely not significant. This is likely due to the greater forecast variability over the assimilation period. On the other hand, a statistically significant improvement in SLP is found in the VORT experiment (approximately 2%) for the same time period. The MOIST and ETS exhibit lower RMSE relative to the CTL, but differences are not statistically significant. These results are consistent with Figures 6 and 7 such that, while the varying flight patterns did not significantly alter the downstream weather, differences up to 8% were found between experiments relative to the CTL forecasts.

Figure 9d–f shows the relative RMSE, similar as for OREGON, for the different experiments relative to the CTL for geopotential height, zonal and meridional winds, RH and temperature at vertical levels of 200–925 hPa. Contrary to the OREGON storm, the ALASKA case shows that the JET experiment consistently reduces RMSE across most variables and pressure levels of approximately 1% to

FIGURE 10 (a, c) Per cent change in total energy error (similar to English *et al.*, 2018) for both the OREGON and ALASKA storm experiments (colors) relative to the control (CTL) experiment as a function of verification date, averaged over the respective Oregon and Alaska verification regions (VRs). Solid lines represent the averages of five forecasts initialized every 6 hr and verified at 2–3 day lead times; dotted lines represent 95% confidence intervals, obtained using a paired *t*-test (two times standard error from the CTL). (b, d) Scatterplots of total energy error (m·s^{-1}) versus the CTL for the OREGON and ALASKA storms at the designated forecast hours relative to the verification time (VT; marker styles) for each respective experiment (colors)
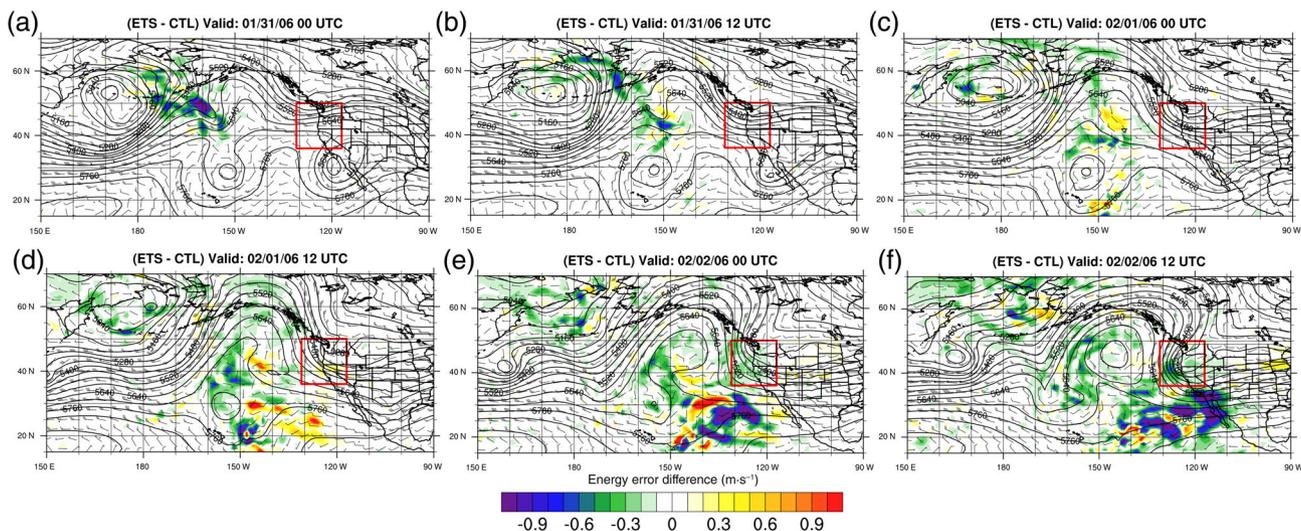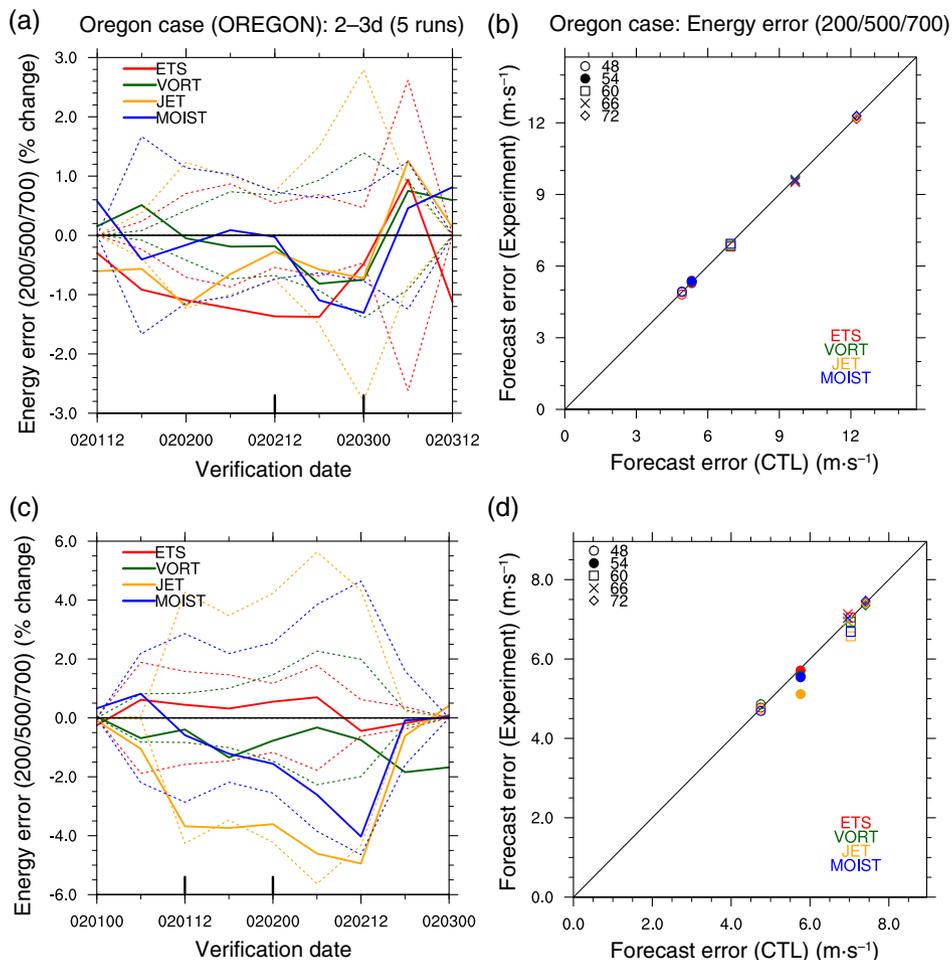
FIGURE 11 Total energy error difference (m·s^{-1}; shaded contours) between the ensemble transform sensitivity (ETS) and control (CTL) experiments (ETS minus the CTL) for the OREGON storm case for a single forecast run initialized on 0000 UTC January 31, 2006. Overlaid are 500 hPa geopotential heights and wind from the nature run (NR). Red boxes denote the verification region as before. Negative differences represent improvement in the forecast in the ETS experiment (reduction in energy error). Differences are shown for (a) initialization time and subsequent forecasts in 12 hr increments (b–f) up to the verification time (VT) at 1200 UTC February 2, 2006
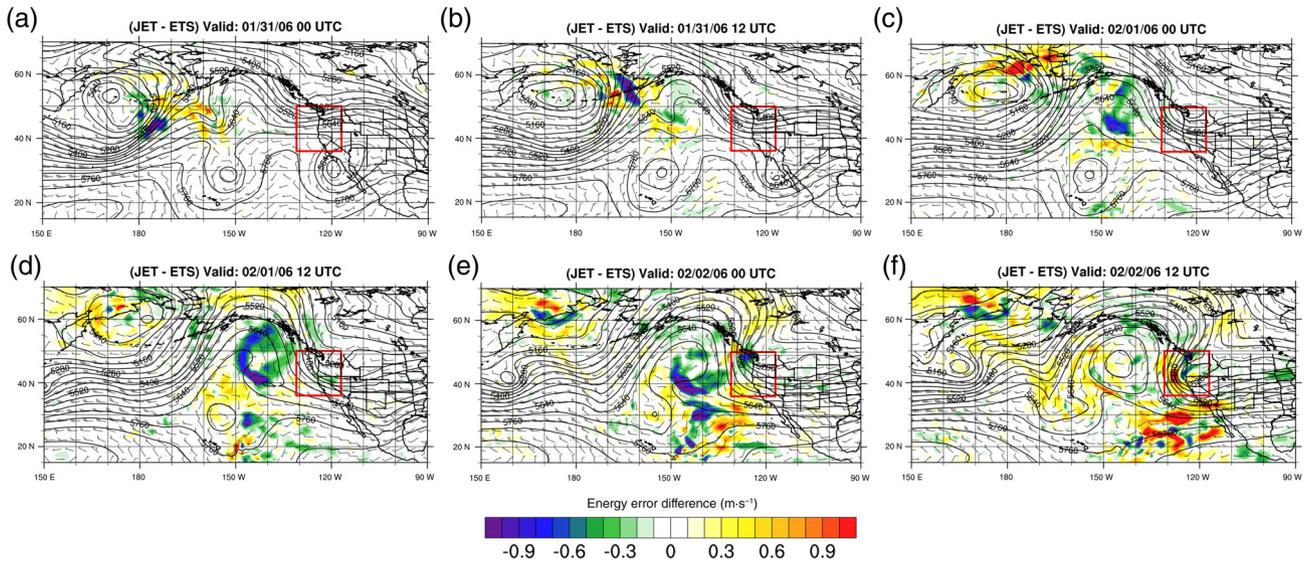
FIGURE 12   As for Figure 11, but for the upper-level jet streaks (JET) experiment minus the ensemble transform sensitivity (ETS)

TABLE 3   Model forecasts at the chosen verification region of Oregon, at a verification time of 1200 UTC February 2, 2006, with reduced error in the OREGON experiments relative to the control (CTL) experiment and relative to the total number of model forecasts. Three metrics are used, similar to English *et al.* (2018): energy error, 500 hPa geopotential height root mean square error (RMSE) and sea-level pressure RMSE

Experiment	Energy error	500 hPa geopotential height	Sea-level pressure
ETS	5/5 (100%)	5/5 (100%)	5/5 (100%)
VORT	2/5 (40%)	3/5 (60%)	2/5 (40%)
JET	4/5 (80%)	2/5 (40%)	4/5 (80%)
MOIST	2/5 (40%)	1/5 (20%)	4/5 (80%)

7%. If one samples a different region, such as with the ETS flight path, the forecast is now largely neutral, further confirming the uncertainty in downstream weather from two different targeting scenarios. These positive forecast benefits are largely found for the upper-level fields, similar to OREGON. Additional positive benefits are found for the VORT and MOIST experiments, but largely in the geopotential height and meridional wind fields (1–2%) (Figure 9d, e). Consistent with the prior storm case, these statistics show that deviations in the flight pattern will lead to sensitivity in weather forecasts downstream. These impacts depend not only on the region under verification but also on the synoptic pattern and the metric and variables used for the evaluation.

4.2.3 | Energy error and propagation of targeted signal into the VR

As in Section 3.2.4, the results are verified in terms of the dry total energy error. Figure 10c shows the *per*

cent change in total energy error relative to the CTL for each dropsonde experiment as a function of verification date over the Alaska VR. Comparable with Figures 6–9, the JET experiment shows the largest reduction in energy error relative to the CTL (Figure 10c, d) at 3–5% between verification dates of 1200 UTC February 1 and 1200 UTC February 2. However, these improvements are not statistically significant. Additional reductions in error are found for VORT and MOIST, although generally smaller than JET and not statistically significant. As noted in Figure 9d–f, the impact of the ETS experiment is largely neutral (differences up to 5%).

Because downstream impacts varied the most between the JET and ETS flight paths, spatial differences are plotted to show how two different flight scenarios can lead to uncertainty in the forecast and, ultimately, different weather impact outcomes.

Figures 19 and 20 show the energy error differences between (1) JET and CTL (Figure 19) and (2) ETS and CTL (Figure 20) for a single forecast
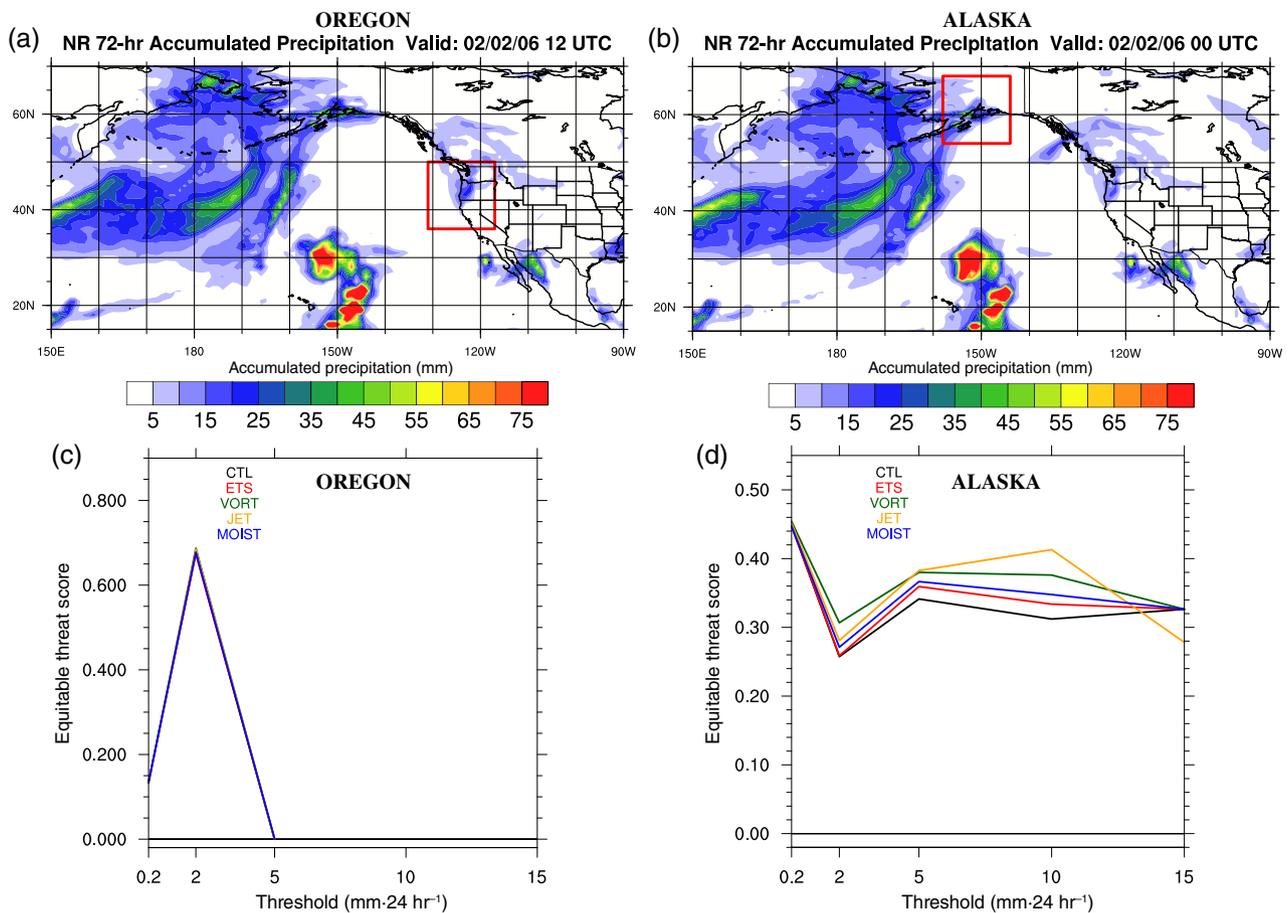
FIGURE 13 (a) Nature run (NR) 72 hr-accumulated precipitation (mm) for the OREGON case from 1200 UTC January 30 to 1200 UTC February 2, 2006. (b) As for (a), but for the 72 hr-accumulated precipitation for the ALASKA case from 0000 UTC January 30 to 0000 UTC February 2, 2006. Red boxes denote the verification region over Oregon and Alaska. Contour interval 5 mm. (c, d) Equitable threat score (EQTS; verified against the NR) for the 24–48 hr forecast period for various precipitation thresholds from 0.2 to 15 mm over a 24 hr period for the control (CTL) experiment and all experiments (colors) for the (c) OREGON and (d) ALASKA cases. Results are averaged over all cycles and over the respective verification regions

initialized at the 2.5 day lead time of 1200 UTC January 30. Both JET and ETS bring the analysis closer to the NR as noted by the negative energy error differences relative to the CTL (Figures 19a and 20a). However, the region with error reduction is found over the western Pacific for JET and the central North Pacific in ETS. The signal at analysis time in both experiments reaches the Alaskan VR by 1200 UTC 31 January (Figures 19c and 20c), despite the fact that the JET flight path samples an area further southwest than the ETS targeting region. While both experiments show improvements in the VR at this time, the JET experiment exhibits larger benefits that persist through 0000 UTC February 2 as opposed to neutral impact in the ETS (Figures 19d–f and 20d–f). Additionally, the JET experiment shows that the area where energy error is reduced relative to the CTL expands to encompass the

entire northeastern Pacific and the western United States, with differences $> 1 \ \mathrm{m \cdot s^{-1}}$ (Figure 19f). This area propagates downstream in connection to the jet stream, consistent with the findings in Irvine et al. (2009). The ETS path, on the other hand, shows mixed results downstream (Figure 20f). Although these results are case dependent, this example clearly shows the uncertainty than can arise from varying flight path designs.

The downstream impacts found for this ALASKA case are generally larger than in the OREGON storm. There are two possible reasons for this. First, the results are likely case dependent. Model physics errors likely play a large role in the outcome of the results, since the analysis error is reduced but the forecast error is not. Second, the analysis errors are larger for the ALASKA case over areas where dropsondes are assimilated in the
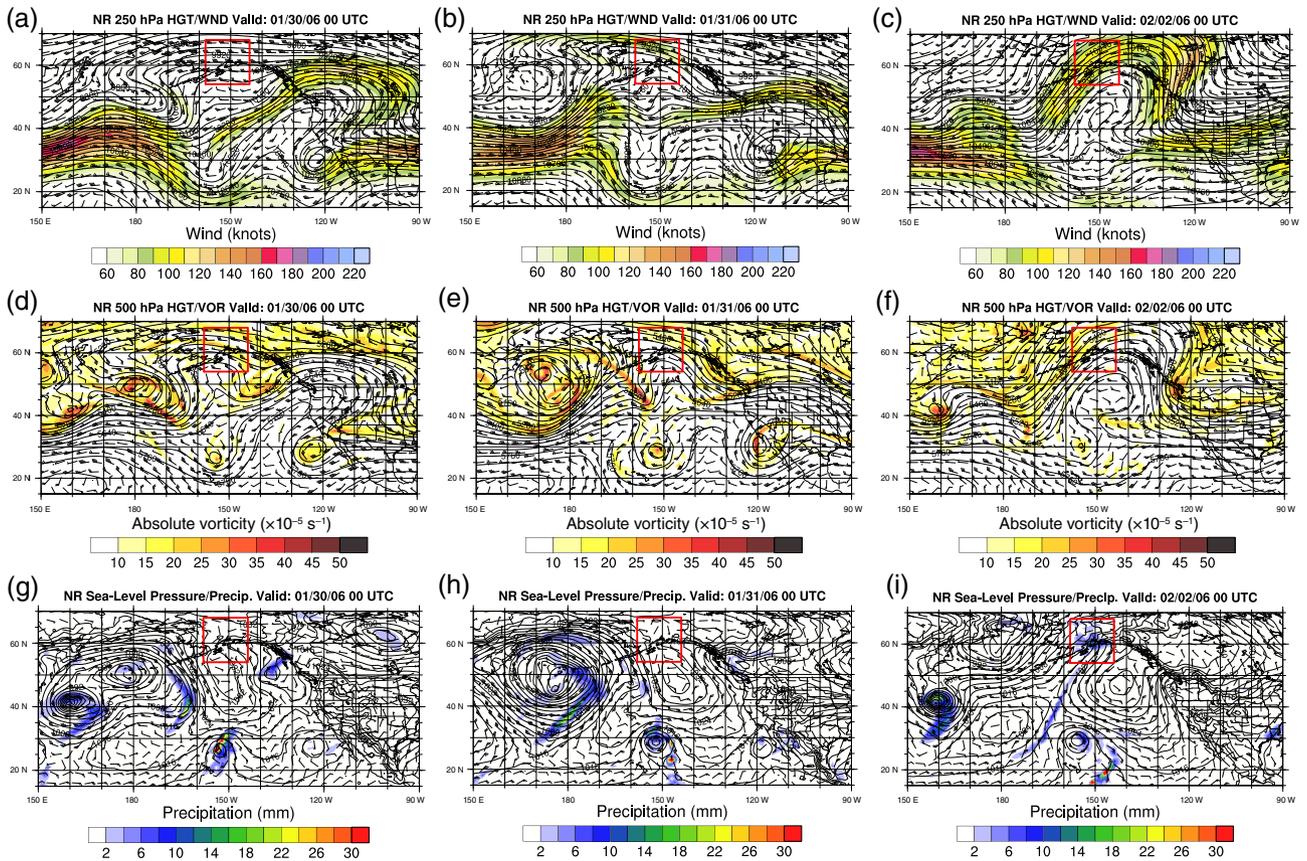
FIGURE 14 As for Figure 1, but for the (a, d, g) 72 hr lead time (0000 UTC January 30, 2006), (b, e, h) 48 hr lead time (0000 UTC January 31, 2006) and (c, f, i) verification time (VT) (0000 UTC February 2, 2006) for the ALASKA observing system simulation experiment (OSSE) storm case. Red boxes denote the verification region Alaska covering a region from 54–68° N to 158–144° W

low to mid-levels of the atmosphere (Figures 2d–i and 15d–i). Therefore, there is more room for improvement in the ALASKA case relative to OREGON. In addition, these analysis errors are largely focused along the storm track in the ALASKA case (Figure 15). It is difficult to state whether these flight track improvements could be reproducible given the small sample sizes of the experiments, but it would be worthwhile to consider in a future study.

Similar to Table 3, Table 4 shows the number of model forecasts at the chosen VR over Alaska at the VT of 0000 UTC February 2, with reduced error relative to the CTL and relative to the total number of model forecasts. Similar to OREGON, the results depend on the flight pattern. The JET and VORT show the largest improvement across energy error, geopotential height and SLP, improving these metrics over CTL 80–100% of the time, while slight changes in the dropsonde locations increase sensitivity and alter impacts for ETS and MOIST, improving forecasts by 20–80%.

4.2.4 | Precipitation verification

Figure 13b shows the accumulated precipitation from the start of model cycling of 0000 UTC January 30 to the VT of 0000 UTC February 2. The ALASKA case contains higher amounts of precipitation in comparison with the OREGON storm, with total precipitation of 45 mm over southern Alaska and near Anchorage, and between 20 and 25 mm over areas of western Alaska. Figure 13d shows the corresponding EQTS, computed over the verification domain 54–68° N and 158–144° W. Unsurprisingly, given the higher precipitations and larger differences in downstream impacts among the four flight patterns, the variability in EQTS is also greater over Alaska. All dropsonde experiments show a higher EQTS than CTL at most precipitation thresholds. The largest improvements are found in the JET and VORT experiments at thresholds up to 10 mm in 24 hr, likely as a result of improving the large-scale synoptic fields. Two reasons for the increased skill in precipitation for the ALASKA case relative to
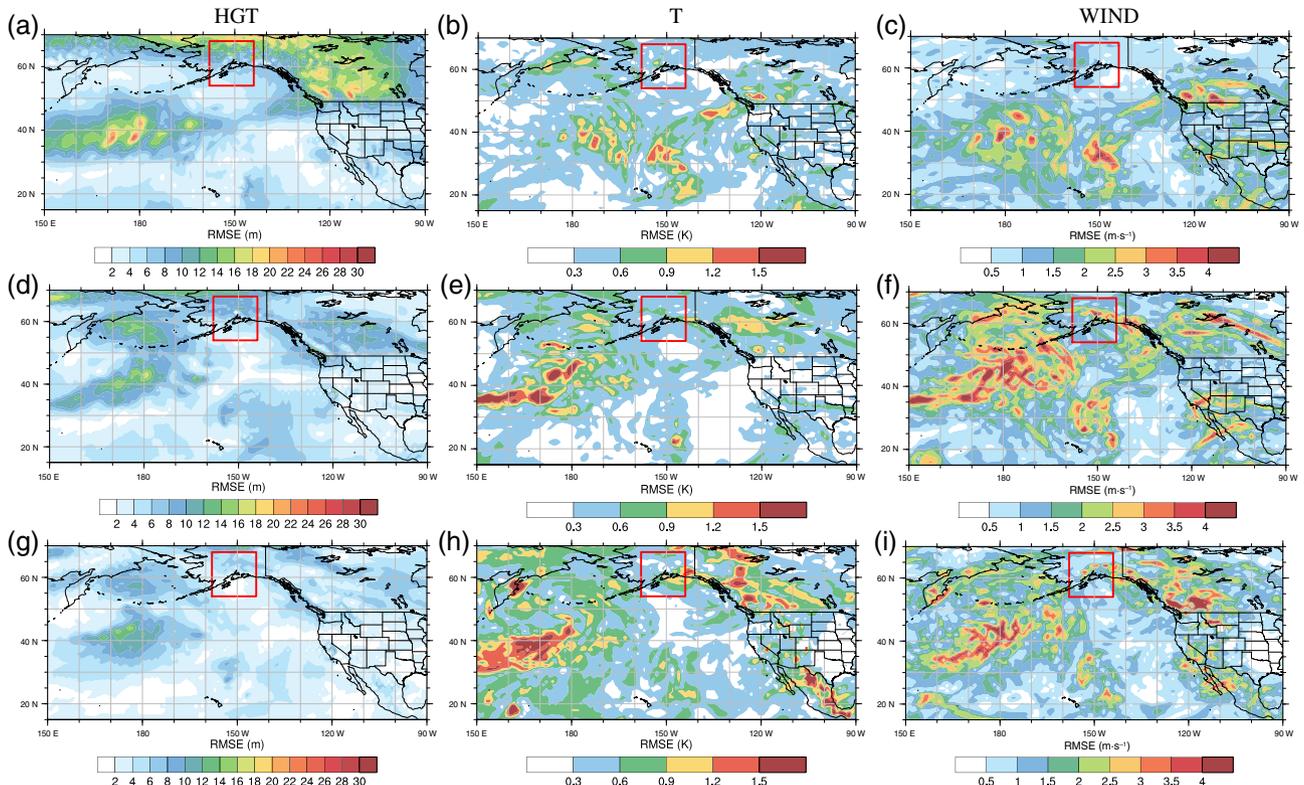
FIGURE 15 As for Figure 2, but for the root mean square error (RMSE) in the control (CTL) experiment for the ALASKA case between 0000 UTC January 30 and 0000 UTC January 31, 2006. Red boxes denote the verification region as in Figure 14

FIGURE 16 As for Figure 3, but for the simulated flight tracks for the ALASKA observing system simulation experiment (OSSE) storm, as in Table 1, from four cycles between 0600 UTC January 28 and 0000 UTC January 29, 2006. All fields are averaged over the 48–72 hr lead time, showing a targeting time of 1200 UTC January 30, 2006, and verification time (VT) of 0000 UTC February 2, 2006. Red dots denote the dropsonde release locations of the Global Hawk (GH). Red boxes denote the verification region as in Figure 14
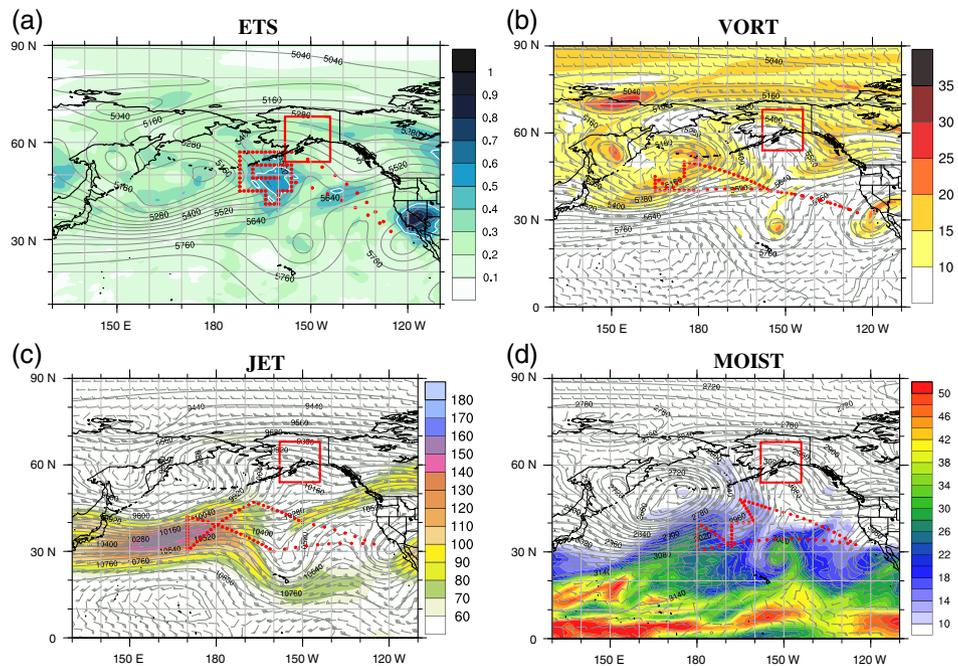

OREGON are likely due to larger high-latitude analysis and forecast errors. These high-latitude errors can arise from the fact that there are fewer observations available at these latitudes. Model physics also contain higher error at these high latitudes. Given a high-impact weather event, changes to the flight path design will impact both the large-scale synoptic fields and precipitation forecasts.
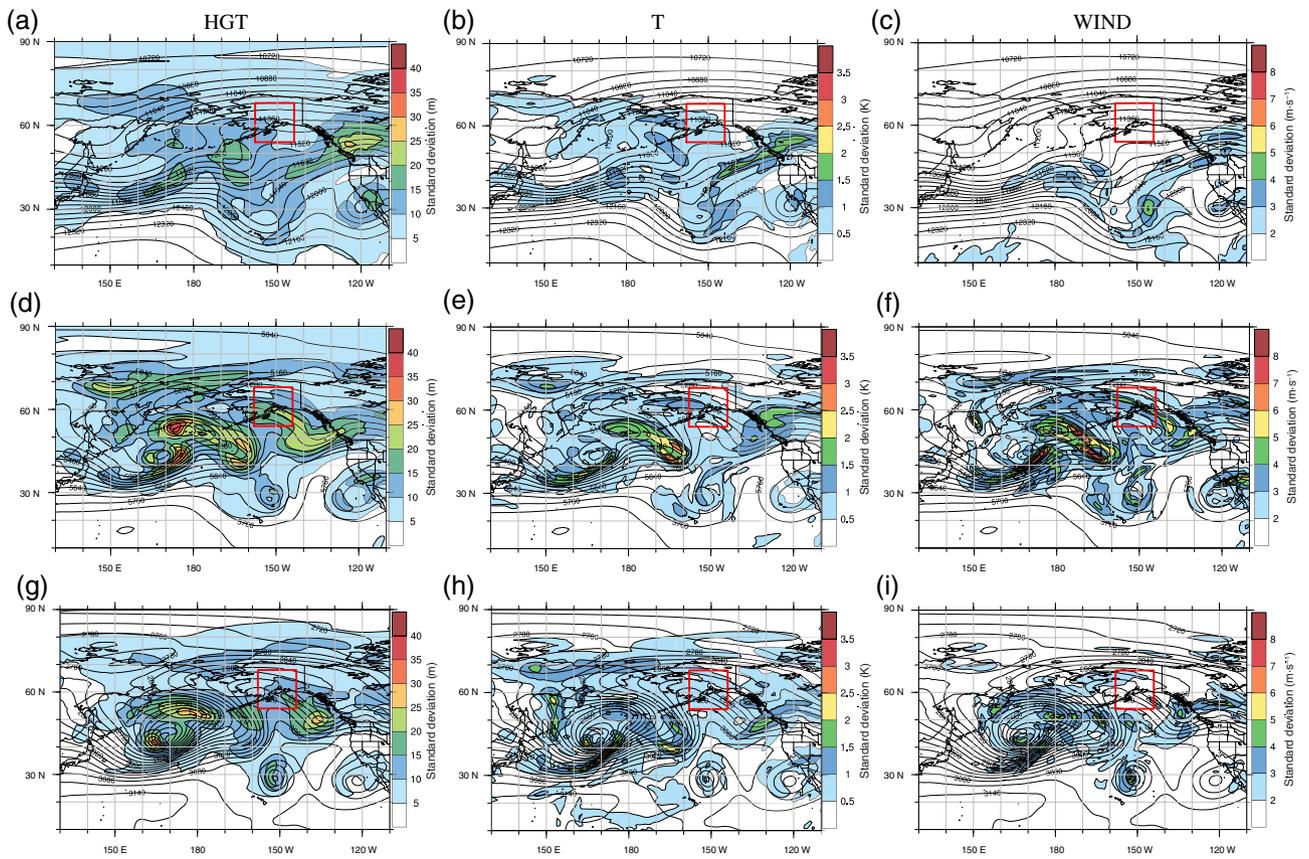
FIGURE 17 As for Figure 4, but for the mean ensemble standard deviation (shaded) and geopotential heights (contour lines) across 80 ensemble members from four cycles between 0600 UTC January 28 and 0000 UTC January 29, 2006. Results are valid at the targeting time of 1200 UTC January 30, 2006, for the ALASKA case. Red boxes denote the verification region as in Figure 14
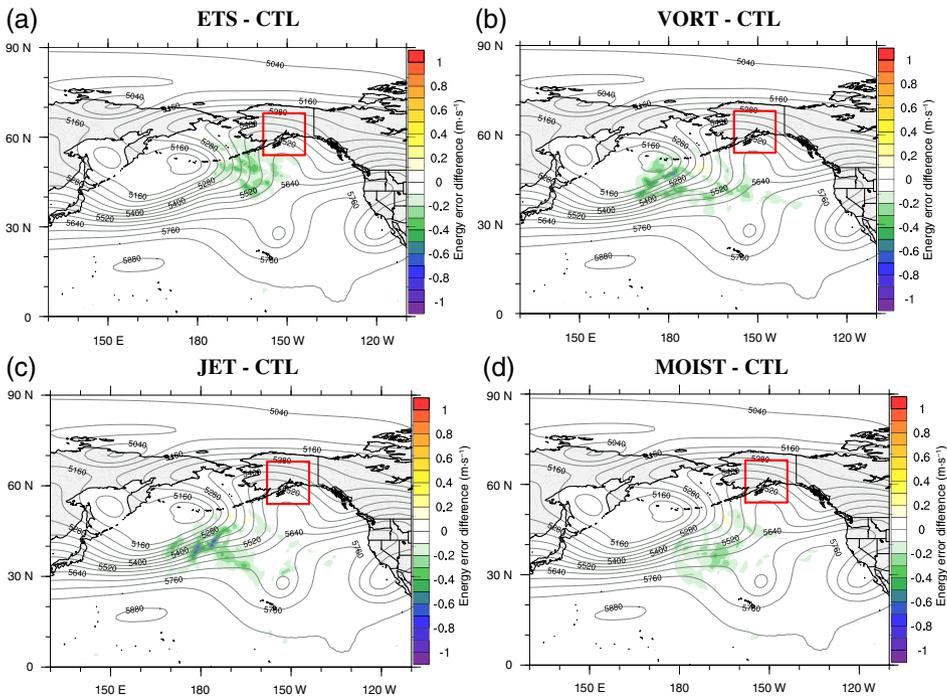
FIGURE 18 As for Figure 5, but for the ALASKA case over the five cycles of dropsonde assimilation (0000 UTC January 30–0000 UTC January 31, 2006). Red boxes denote the verification region as in Figure 14
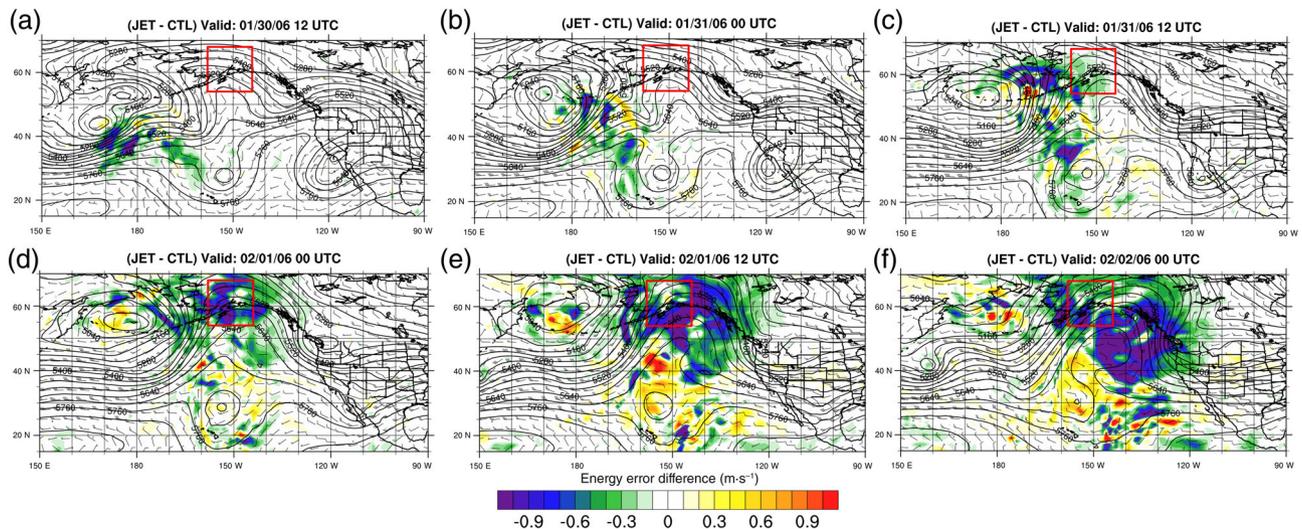
FIGURE 19 As for Figure 11, but for the ALASKA observing system simulation experiment (OSSE) storm for a single forecast run initialized on 1200 UTC January 30, 2006, for the upper-level jet streaks (JET) minus the control (CTL) experiment. Red boxes denote the verification region over as in Figure 14
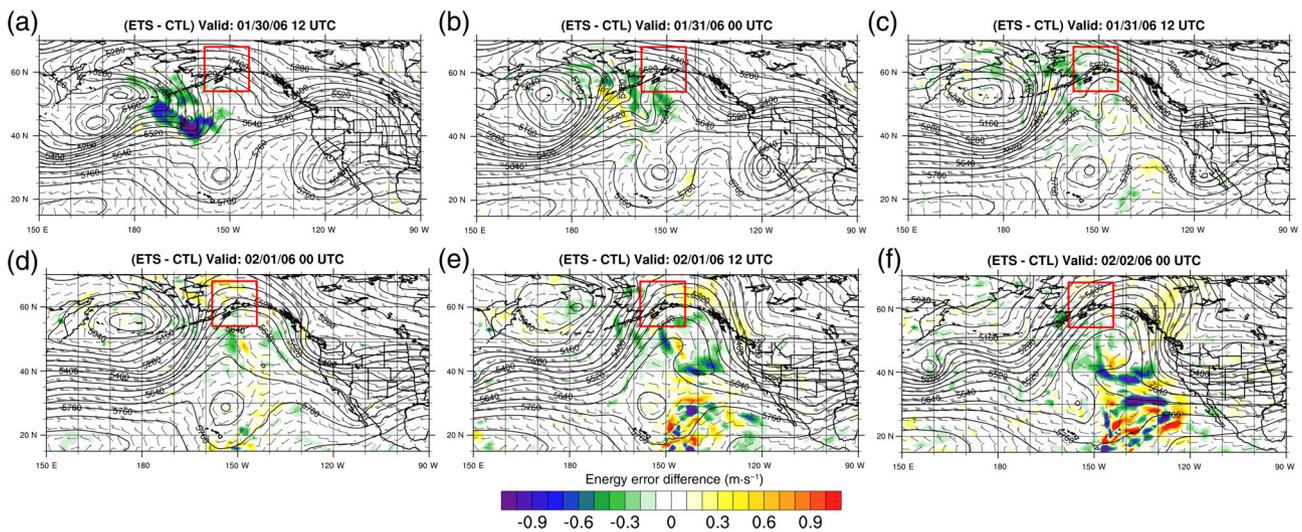
FIGURE 20 As for Figure 19, but for the ensemble transform sensitivity (ETS) experiment minus the control (CTL) experiment

TABLE 4 As for Table 3, but for the ALASKA experiments at a verification time of 0000 UTC February 2, 2006

Experiment	Energy error	500 hPa geopotential height	Sea-level pressure
ETS	1/5 (20%)	4/5 (80%)	4/5 (80%)
VORT	4/5 (80%)	4/5 (80%)	5/5 (100%)
JET	4/5 (80%)	5/5 (100%)	5/5 (100%)
MOIST	3/5 (60%)	3/5 (60%)	3/5 (60%)

5 | DISCUSSION AND CONCLUSIONS

The study has investigated the sensitivity (or uncertainty) of forecast impact to operational implementation and design of flight patterns for downstream winter weather forecasts. A preliminary investigation was performed with varying flight tracks using NASA's Global Hawk (GH) platform in an observing system simulation experiment (OSSE) framework. Eight flight patterns were

created in a manner mimicking what could occur in an operational field campaign. These experiments were carried out across two winter storms and the results were compared by analysing the downstream impacts.

For both storm cases, downstream forecasts were found to be sensitive to the orientation and pattern of the flight track over the targeting region. Some flight patterns resulted in significant improvements over the predefined VRs and/or downstream over the contiguous United States (CONUS). Uncertainty among different flight paths ranged between approximately 1% and 8% depending on the metric used and the atmospheric variables analysed.

While all flight paths improved the analysis, the propagation of the signal from the targeting region to the VR was found to be highly case dependent. This was most evident for the ALASKA storm, in which the JET experiment showed large positive improvements in the verification region (VR) and downstream over the eastern Pacific and CONUS, whereas the ensemble transform sensitivity (ETS) path showed a neutral impact. Conversely, for the OREGON storm, the ETS path showed a statistically significant improvement of 1–2%. These findings show that changes to the proposed flight path play an important role in weather forecasting and the impact varies on a case-by-case basis. This information could be used by mission scientists and managers when planning flight mission designs.

The use of a time-averaged ETS signal dampened the sensitivity in certain regions, specifically the signal associated with faster propagating synoptic systems. These regions also tended to occur in locations where analysis errors grew rapidly. The use of an 80-member ensemble every 6 hr instead of a lagged 20-member ensemble may also provide more accurate error variance estimates at the targeting time, as well as error covariance structures between the targeting and verification times (VTs) in the ETS technique. In addition, the upper-level signal could be weakened if only three vertical levels are used in the ETS method. Finally, other error norms such as variance in sea level pressure and accumulated precipitation could be used to verify weather forecasts. As in previous studies, it is shown that flight patterns should sample an area upstream as large as possible (e.g. Majumdar *et al.*, 2002a, 2002b).

The study only considered two storms and used perfect simulated observations in the OSSEs. Future studies should test more cases with different flight scenarios over a much larger composite of storms and using realistic observation errors. By incorporating several different atmospheric features (tropical and extratropical cyclones, atmospheric rivers, blocking patterns, mid-level ridges and jet streaks, and so on) with targeted guidance over several cases, it may be possible to develop a composite of flight tracks that could be used for a variety of weather scenarios. These proposed flight tracks could then be tested in an OSSE environment, helping to provide a more actionable strategy to reduce forecast uncertainty caused by flight track design.

ORCID

A. C. Kren https://orcid.org/0000-0002-8404-3096

REFERENCES

Aberson, S.D., Sellwood, K.J. and Leighton, P.A. (2017) Calculating Dropwindsonde location and time from TEMP-DROP messages for accurate assimilation and analysis. *Journal of Atmospheric and Oceanic Technology*, 34, 1673–1678. https://doi.org/10.1175/JTECH-D-17-0023.1.

Andersson, E. and Masutani, M. (2010) Collaboration on observing system simulation experiments (joint OSSE). *Bulletin of the American Meteorological Society*, 67(6), 687–695.

Atlas, R., Hoffman, R.N., Leidner, S.M., Sienkiewicz, J., Yu, T.-W., Bloom, S.C., Brin, E., Ardizzone, J., Terry, J., Bungato, D. and Jusem, J.C. (2001) The effects of marine winds from Scatterometer data on weather analysis and forecasting. *Bulletin of the American Meteorological Society*, 82(9), 1965–1990. https://doi.org/10.1175/1520-0477.

Atlas, R., Hoffman, R.N., Ma, Z., Emmitt, G.D., Wood, S.A., Jr., Greco, S., Tucker, S., Bucci, L., Annane, B., Hardesty, R.M. and Murillo, S. (2015) Observing system simulation experiments (OSSEs) to evaluate the potential impact of an optical autocovariance wind lidar (OAWL) on numerical weather prediction. *Journal of Atmospheric and Oceanic Technology*, 32, 1593–1613. https://doi.org/10.1175/JTECH-D-15-0038.1.

Black, M., Wick, G.A. and Hood, R.E. (2014) Sensing hazards with operational unmanned technology: NOAA's multi-year plan to deploy the NASA Global Hawk aircraft for high impact weather. 31st Conference on Hurricanes and Tropical Meteorology, San Diego, CA, American Meteorological Society, 5A.7 [Cited November 1, 2019]. Available at: https://ams.confex.com/ams/31Hurr/webprogram/Paper244777.html [Accessed May 1, 2017].

Campins, J., Navascués, B., Santos, C. and Amo-Baladrón, A. (2013) Influence of targeted observations on short-term forecasts of high-impact weather events in the Mediterranean. *Natural Hazards and Earth System Sciences*, 13, 2891–2910. https://doi.org/10.5194/nhess-13-2891-2013.

Christophersen, H., Aksoy, A., Dunion, J. and Sellwood, K. (2017) The impact of NASA global hawk unmanned aircraft Dropwindsonde observations on tropical cyclone track, intensity, and structure: case studies. *Monthly Weather Review*, 145, 1817–1830. https://doi.org/10.1175/MWR-D-16-0332.1.

Christophersen, H., Aksoy, A., Dunion, J. and Aberson, S. (2018a) Composite impact of global hawk unmanned aircraft dropwindsondes on tropical cyclone analyses and forecasts. *Monthly Weather Review*, 146, 2297–2314. https://doi.org/10.1175/MWR-D-17-0304.1.

Christophersen, H., Atlas, R., Aksoy, A. and Dunion, J. (2018b) Combined use of satellite observations and global hawk unmanned aircraft dropwindsondes for improved tropical cyclone analyses and forecasts. *Weather Forecasting*, 33, 1021–1031. https://doi.org/10.1175/WAF-D-17-0167.1.

Cucurull, L., Li, R. and Peevey, T.R. (2017) Assessment of radio occultation observations from the COSMIC-2 Mission with a simplified observing system simulation experiment configuration. *Monthly Weather Review*, 145, 3581–3597. https://doi.org/10.1175/MWR-D-16-0475.1.

Dole, R.M., Spackman, J.R., Newman, M., Compo, G.P., Smith, C.A., Hartten, L.M., Barsugli, J.J., Webb, R.S., Hoerling, M.P., Cifelli, R., Wolter, K., Barnet, C.D., Gehne, M., Gelaro, R., Kiladis, G.N., Abbott, S., Akish, E., Albers, J., Brown, J.M., Cox, C.J., Darby, L., De Boer, G., DeLuisi, B., Dias, J., Dunion, J., Eischeid, J., Fairall, C., Gambacorta, A., Gorton, B.K., Hoell, A., Intrieri, J., Jackson, D., Johnston, P.E., Lataitis, R., Mahoney, K.M., McCaffrey, K., Mueller, M.J., Murray, D., Neiman, P.J., Otto, W., Persson, O., Quan, X.-W., Rangwala, I., Ray, A.J., Reynolds, D., Dellaripa, E.R., Rosenlof, K., Sakaeda, N., Sardeshmukh, P.D., Slivinski, L.C., Smith, L., Solomon, A., Swales, D., Tulich, S., White, A., Wick, G., Winterkorn, M.G., Wolfe, D.E. and Zamora, R. (2018) Advancing science and services during the 2015–16 El Niño: the NOAA El Niño rapid response field campaign. *Bulletin of the American Meteorological Society*, 99, 975–1001. https://doi.org/10.1175/BAMS-D-16-0219.1.

Dunion, J.P., Wick, G.A., Black, P.G. and Walker, J. (2018) Sensing hazards with Operational Unmanned Technology: 2015–2016 Campaign Summary, Final Report. NOAA Tech Memo. OAR-UAS-001, 49 pp. [Cited November 1, 2019]. Available at: https://doi.org/10.7289/V5/TM-OAR-UAS-001 [Accessed 1 January 2018].

English, J.M., Kren, A.C. and Peevey, T.R. (2018) Improving winter storm forecasts with observing system simulation experiments (OSSEs). Part 2: Evaluating a Satellite Gap with Idealized and Targeted Dropsondes. *Earth and Space Science*, 5, 176–196. https://doi.org/10.1002/2017EA000350-T.

Errico, R., Yang, R., Privé, N., Tai, K.-T., Todling, R., Sienkiewicz, M. and Guo, J. (2013) Development and validation of observing-system simulation experiments at NASA's global modeling and assimilation office. *Quarterly Journal of the Royal Meteorological Society*, 139, 1162–1178. https://doi.org/10.1002/qj.2027.

Halliwell, G.R., Srinivasan, A., Kourafalou, V., Yang, H., Willey, D., Le Henaff, M. and Atlas, R. (2014) Rigorous evaluation of a fraternal Twin Ocean OSSE system for the open Gulf of Mexico. *Journal of Atmospheric and Oceanic Technology*, 31, 105–130. https://doi.org/10.1175/JTECH-D-13-00011.1.

Hamill, T.M., Yang, F., Cardinali, C. and Majumdar, S.J. (2013) Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Monthly Weather Review*, 141, 2058–2065. https://doi.org/10.1175/MWR-D-12-00309.1.

Hoffman, R.N. and Atlas, R. (2016) Future observing system simulation experiments. *Bulletin of the American Meteorological Society*, 97, 1601–1616. https://doi.org/10.1175/BAMS-D-15-00200.1.

Irvine, E.A., Gray, S.L., Methven, J., Renfrew, I.A., Bovis, K. and Swinbank, R. (2009) The impact of targeted observations made during the Greenland flow distortion experiment. *Quarterly Journal of the Royal Meteorological Society*, 135, 2012–2029. https://doi.org/10.1002/qj.499.

Kleist, D.T. and Ide, K. (2015a) An OSSE-based evaluation of hybrid Variational-ensemble data assimilation for the NCEP GFS. Part I: system description and 3D-hybrid results. *Monthly Weather Review*, 143, 433–451. https://doi.org/10.1175/MWR-D-13-00351.1.

Kleist, D.T. and Ide, K. (2015b) An OSSE-based evaluation of hybrid Variational-ensemble data assimilation for the NCEP GFS. Part II: 4DEnVar and hybrid variants. *Monthly Weather Review*, 143, 452–470. https://doi.org/10.1175/MWR-D-13-00350.1.

Kren, A.C., Wang, H., English, J.M., Peevey, T.R. and Cucurull, L. (2016) Impact of Global Hawk Data from the SHOUT-ENRR 2016 Field Campaign on an Atmospheric River in the Central North Pacific. 1st International Atmospheric Rivers Conference (IARC), La Jolla, CA [Cited November 1, 2019]. Available at: http://cw3e.ucsd.edu/ARconf2016/Program_Presentations/. [Accessed 1 November 2017].

Kren, A.C., Cucurull, L. and Wang, H. (2018) Impact of UAS global hawk dropsonde data on tropical and extratropical cyclone forecasts in 2016. *Weather Forecasting*, 33, 1121–1141. https://doi.org/10.1175/WAF-D-18-0029.1.

Langland, R.H. (2005) Issues in targeted observing. *Quarterly Journal of the Royal Meteorological Society*, 131, 3409–3425. https://doi.org/10.1256/qj.05.130.

Leidner, S.M., Nehrkorn, T., Henderson, J. and Mountain, M. (2017) A severe weather quick observing system simulation experiment (QuickOSSE) of global navigation satellite system (GNSS) radio occultation (RO) Superconstellations. *Monthly Weather Review*, 145, 637–651. https://doi.org/10.1175/JTECH-D-13-00011.1.

Majumdar, S. (2016) A review of targeted observations. *Bulletin of the American Meteorological Society*, 97, 2287–2303. https://doi.org/10.1175/BAMS-D-14-00259.1.

Majumdar, S.J., Bishop, C.H., Buizza, R. and Gelaro, R. (2002a) A comparison of ensemble-transform Kalman-filter targeting guidance with ECMWF and NRL total-energy singular-vector guidance. *Quarterly Journal of the Royal Meteorological Society*, 128, 2527–2549. https://doi.org/10.1256/qj.01.214.

Majumdar, S.J., Bishop, C.H., Etherton, B.J. and Toth, Z. (2002b) Adaptive sampling with the ensemble transform Kalman filter. Part II: Field program implementation. *Monthly Weather Review*, 130, 1356–1369. https://doi.org/10.1175/1520-0493.

Majumdar, S.J., Sellwood, K.J., Hodyss, D., Toth, Z. and Song, Y. (2010) Characteristics of target areas selected by the ensemble transform Kalman filter for medium-range forecasts of high-impact winter weather. *Monthly Weather Review*, 138, 2803–2824. https://doi.org/10.1175/2010MWR3106.1.

Masutani, M. Andersson, E., Terry, J., Reale, O., Jusem, J.C., Riishojgaard, L.P., Schlatter, T., Stoffelen, A., Woollen, J., Lord, S., Toth, Z., Song, Y., Dleist, D., Xie, Y., Prive, N., Liu, E., Sun, H., Emmitt, D., Greco, S., Wood, S.A., Marseille, G.-J., Errico, R., Yang, R., McConaughy, G., Devenyi, D., Weygandt, S., Tompkins, A., Jung, T., Anantharaj, V., Hill, C., Fitzpatrick, P., Weng, F., Zhu, T., Boukabara, S. (2007) 18th Conference on Numerical Weather Prediction, Park City, UT, American Meteorological Society, 12B.5 [Cited November 1, 2019]. Available at: http://ams.confex.com/ams/pdfpapers/124080.pdf [Accessed September 8, 2018].

McCarty, W., Errico, R. and Gelaro, R. (2012) Cloud coverage in the joint OSSE nature run. *Monthly Weather Review*, 140(6), 1863–1871. https://doi.org/10.1175/MWR-D-11-00131.1.

McClung, T. (2014) Technical Implementation Notice 14–46. Washington, DC, USA, Office of Science and Technology [Cited November 1, 2019]. Available at: http://www.nws.noaa.gov/om/notification/tin14-46gfs.htm [Accessed May 1, 2017].

McNally, A.P. (2002) A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Quarterly Journal of the Royal Meteorological Society*, 128, 2551–2556. https://doi.org/10.1256/qj.01.206.

McNally, A.P. (2009) The direct assimilation of cloud-affected satellite infrared radiances in the ECMWF 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 135, 1214–1229. https://doi.org/10.1002/qj.426.

Peevey, T.R., English, J.M., Cucurull, L., Wang, H. and Kren, A.C. (2018) Improving winter storm forecasts with observing system simulation experiments (OSSEs): part 1, an idealized case study of three US storms. *Monthly Weather Review*, 146, 1341–1366. https://doi.org/10.1175/MWR-D-17-0160.1.

Privé, N.C., Xie, Y., Koch, S., Atlas, R., Majumdar, S.J. and Hoffman, R.N. (2014) An observing system simulation experiment for the unmanned aircraft system data impact on tropical cyclone track forecasts. *Monthly Weather Review*, 142, 4357–4363. https://doi.org/10.1175/MWR-D-14-00197.1.

Ralph, F.M., Neiman, P.J. and Rotunno, R. (2005) Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-1998 and PACJET-2001: mean vertical profile and atmospheric-river characteristics. *Monthly Weather Review*, 133, 889–910. https://doi.org/10.1175/MWR2896.1.

Reale, O., Terry, J., Masutani, M., Andersson, E., Riishojgaard, L. and Jusem, J.C. (2007) Preliminary evaluation of the European Centre for Medium-range Weather Forecasts (ECMWF) nature run over the tropical Atlantic and African monsoon region. *Geophysical Research Letters*, 34, L22810.

Ryan, K.E., Bucci, L.R., Christophersen, H., Atlas, R.M., Murillo, S. and Dodge, P. (2015) OSSE evaluation of prospective aircraft reconnaissance flight patterns and their impact on Hurricane forecasts. American Geophysical Union – Fall Meeting, San Francisco, CA; A41E-0110.

Ryan, K., Bucci, L., Atlas, R., Delgado, J. and Murillo, S. (2019) Impact of gulfstream-IV Dropsondes on tropical cyclone prediction in a regional OSSE system. *Monthly Weather Review*, 147, 2961–2977. https://doi.org/10.1175/MWR-D-18-0157.1.

Sellwood, K.J., Majumdar, S.J., Mapes, B.E. and Szunyogh, I. (2008) Predicting the influence of observations on medium-range forecasts of atmospheric flow. *Quarterly Journal of the Royal Meteorological Society*, 134, 2011–2027. https://doi.org/10.1002/qj.341.

Sippel, J.A., Friedman, K.L. and Tallapragada, V. (2018) *The impact of Global Hawk dropsondes on NCEP hurricane forecasts*. 33rd Conference on Hurricanes and Tropical Meteorology, Ponte Vedra, FL, American Meteorological Society, 8D.6 [Cited November 1, 2019]. Available at: http://ams.confex.com/ams/33HURRICANE/webprogram/Paper338883.html [Accessed 8 September 2018].

Szunyogh, I., Toth, Z., Morss, R.E., Majumdar, S.J., Etherton, B.J. and Bishop, C.H. (2000) The effect of targeted Dropsonde observations during the 1999 winter storm reconnaissance program. *Monthly Weather Review*, 128, 3520–3537. https://doi.org/10.1175/1520-0493.

Szunyogh, I., Toth, Z., Zimin, A.V., Majumdar, S.J. and Persson, A. (2002) Propagation of the effect of targeted observations: the 2000 winter storm reconnaissance program. *Monthly Weather Review*, 130, 1144–1165. https://doi.org/10.1175/1520-0493.

Wang, H., Xie, Y., Peter, C., Fan, S.-Y. and Zhang, Y. (2018) Ensemble transform sensitivity for adaptive observations: a general formulation and its practicable implementation. *Atmospheric and Oceanic Science Letters*, 11, 352–357. https://doi.org/10.1080/16742834.2018.1491778.

Wick, G., Dunion, J. and Walker, J. (2018) Sensing hazards with operational unmanned technology: Impact study of Global Hawk unmanned aircraft system observations for hurricane forecasting, final report. NOAA Tech Memo. OAR-UAS-002, 92 pp. [Cited November 1, 2019]. Available at: ftp://ftp1.esrl.noaa.gov/et6/sat/shout/data_impact/TM_SHOUT_Impact_Assessment_FINAL.pdf [Accessed 5 Mar 2019].

Wick, G., Dunion, J.P., Black, P.G., Walker, J.R., Torn, R.D., Kren, A.C., Aksoy, A., Christophersen, H., Cucurull, L., Dahl, B., English, J.M., Friedman, K., Peevey, T.R., Sellwood, K., Sippel, J.A., Tallapragada, V., Taylor, J., Wang, H., Hood, R.E. and Hall, P. (2020) NOAA's sensing hazards with operational unmanned technology (SHOUT) experiment: observations and forecast impacts. *Bulletin of the American Meteorological Society*, 101, E968–E987. https://doi.org/10.1175/BAMS-D-18-0257.1.

Wilks, D.S. (2006) *Statistical Methods in the Atmospheric Sciences*. Burlington, MA: Academic Press.

Winterbottom, Henry. Research Scientist. (Personal Communication, 1st February 2019).

Zhang, Y., Xie, Y.F., Wang, H.L., Chen, D.H. and Toth, Z. (2016) Ensemble transform sensitivity method for adaptive observations. *Advances in Atmospheric Sciences*, 33(1), 10–20. https://doi.org/10.1007/s00376-015-5031-9.

Zhu, T., Weng, F., Masutani, M. and Woollen, J.S. (2012) Synthetic radiance simulation and evaluation for a joint observing system simulation experiment. *Journal of Geophysical Research*, 117, D23111. https://doi.org/10.1029/2012JD017697.