NOAA Technical Memorandum OAR UAS-002

Sensing Hazards with Operational Unmanned Technology: Impact Study of Global Hawk Unmanned Aircraft System Observations for Hurricane Forecasting, Final Report

Gary A. Wick
Jason P. Dunion
John Walker

NOAA/Unmanned Aircraft Systems Program
Silver Spring, Maryland
January 2018

noaa NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION / Office of Oceanic and Atmospheric Research

NOAA Technical Memorandum OAR UAS-002

https://doi.org/10.7289/V5/TM-OAR-UAS-002

Sensing Hazards with Operational Unmanned Technology: Impact Study of Global Hawk Unmanned Aircraft System Observations for Hurricane Forecasting, Final Report

Gary A. Wick
Jason P. Dunion
John Walker

NOAA/Unmanned Aircraft Systems
Silver Spring, Maryland

January 2018

NOTICE

This document was prepared as an account of work sponsored by an agency of the United States Government. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency or Contractor thereof. Neither the United States Government, nor Contractor, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, product, or process disclosed, or represents that its use would not infringe privately owned rights. Mention of a commercial company or product does not constitute an endorsement by the National Oceanic and Atmospheric Administration Office of Oceanic and Atmospheric Research. Use of information from this publication concerning proprietary products or the tests of such products for publicity or advertising purposes is not authorized.

CONTRIBUTORS TO CITED ANALYSES

Altug Aksoy
Robert Atlas
Hui Christophersen
Lidia Cucurull
Brittany Dahl
James Doyle
Jason English
Kate Howard
Andrew Kren
Tanya Peevey
Nikki Privé
Kathryn Sellwood
Jason Sippel
Vijay Tallapragada
James Taylor
Ryan Torn
Zoltan Toth
Hongli Wang
Yuanfu Xie

EXECUTIVE SUMMARY

The primary scientific goal of the Sensing Hazards with Operational Unmanned Technology (SHOUT) Project is to determine the potential utility of observations from high-altitude, long-endurance unmanned aircraft systems (UAS) such as the Global Hawk (GH) aircraft to improve forecasts of high-impact weather events or mitigate potential degradation of forecasts in the event of a future gap in satellite coverage. Hurricanes and tropical cyclones are among the most potentially destructive high-impact weather events and pose a major forecasting challenge to NOAA. Major winter storms over the Pacific Ocean, including atmospheric river events, which make landfall and bring strong winds and extreme precipitation to the West Coast and Alaska are also important to forecast accurately because of their societal impact in those parts of the country. In response, the SHOUT project has supported field campaigns with the GH aircraft and dedicated data impact studies exploring the potential to improve the forecasting of both tropical cyclones and landfalling Pacific storms.

Three major field campaigns were led by the SHOUT project to collect in situ and remote sensing observations to support the data impact assessments. Details on these campaigns, the deployed sensors, and additional collaboration with experiments led by other agencies is described in a companion campaign summary document (Dunion et al. 2018). This present document is the final report summarizing the scientific results of the multiple studies evaluating the potential forecast impact of GH observations conducted under or in collaboration with the SHOUT project. This final report supersedes interim impact assessment reports prepared in 2014-2016. The SHOUT project also analyzed the cost and operational effectiveness of using the GH aircraft. Those results are presented in the companion cost assessment report (Kenul et al. 2018).

The data impact assessments incorporate both data denial studies using actual GH data, termed Observing System Experiments (OSEs), and Observing System Simulation Experiments (OSSEs) based on simulated observations taken from a "nature" run assumed to represent reality. SHOUT funded analyses were conducted by teams at the Hurricane Research Division (HRD) of the Atlantic Oceanographic and Meteorological Laboratory (AOML) and the Global Systems Division (GSD) of the Earth System Research Laboratory (ESRL). Additional collaborative analyses were performed in partnership with SHOUT at the Environmental Modeling Center within NOAA's National Centers for Environmental Prediction (NCEP/EMC). The studies have evaluated the effects of GH-type observations on forecasts of tropical storms and Pacific winter storms using both global and regional models.

Assessment of the potential impact of GH observations on regional scale tropical cyclone forecasts was based on use of the Hurricane Weather Research and Forecasting (HWRF) model. Studies employed the current operational model configuration, a previous operational version to explicitly evaluate the value of GH observations in the event of a satellite gap, and a

version incorporating the Hurricane Ensemble Data Assimilation (HEDAS) system from HRD. Evaluations with the 2017 operational version still underway at EMC for storms sampled in 2016 show notable positive improvements in both track and intensity forecasts at longer lead times resulting from assimilation of GH dropsonde data. Skill improvements reach up to about 9% for track and 14% for intensity. Tests using the version of the model operational through mid-2016 were conducted for Hurricane Matthew (2016) for the scenario where sounding observations from the Suomi National Polar-orbiting Partnership (NPP) satellite were withheld to simulate a gap in satellite coverage. The results showed positive track forecast impacts from assimilation of both GH dropsonde observations and remote microwave sounder retrievals of temperature and humidity, especially for model cycles where the observations were directly assimilated. Track forecast skill improvements approached 30% while intensity impacts were more mixed.

Experiments using HEDAS were conducted for a composite study of the impact of dropsonde observations from ten storm systems and for initial analyses of the impact of other sensor types for subsets of the storms. The composite investigation showed a large positive impact of the dropsondes on the analyzed storm structure in excess of 25% and a corresponding increase in the skill of track and intensity forecasts frequently exceeding 10%. The intensity improvements were largely limited to cases where no other aircraft reconnaissance data were available. Preliminary experiments assimilating microwave and infrared sounder retrieved profiles and microwave-derived surface wind speed fields all exhibited periods of positive forecast impact, but more work is still required to obtain conclusive results. The GH observations were also observed to be a powerful supplement to satellite measurements.

The impact of GH dropsonde observations on forecasts from the operational Global Forecast System (GFS) model was examined in detail. Experiments focused on tropical cyclones were conducted using the new 2017 operational version at EMC for all observed storms in 2014 and 2016 in the presence of the full conventional observing system, and using the 2015 version at GSD for Hurricane Matthew to compare the impact with and without a gap in sounding data from the Suomi NPP satellite. All the experiments demonstrated notable positive improvements in the track forecast at longer lead times. The results from EMC were particularly significant, demonstrating multi-storm average track skill improvements exceeding 10% and improvements for individual storms of over 20% depending on forecast lead time. The results also showed improvements in the track forecasts of concurrent Pacific cyclones based on observations of the Atlantic storms, suggesting that the observations could have positive larger-scale impacts. The GSD results showed notable positive improvements in forecasted precipitation off the southeastern coast during Hurricane Matthew resulting from the track improvements. Positive impacts were observed in multiple forecast fields both with and without a gap in satellite coverage. The similarity in results across different variables suggested the results were largely independent of the selected metric. The positive benefits were generally greater in the event of a satellite gap.

An overall summary capturing the broad potential for GH dropsonde observations to significantly improve tropical cyclone forecast accuracy across research and operational modeling systems is presented in Table ES-1. The table documents the percent improvement in forecasted track accuracy for hurricanes and tropical storms at a specific 96-hr lead time for the multiple model configurations and scenarios evaluated. All systems tested demonstrated positive forecast impacts, particularly for data from those flights focused on Hurricane Matthew in 2016. The potential for overall gains of 12% and gains in excess of 25% for individual systems within the current operational modeling systems is remarkable. The corresponding forecast intensity was also improved by up to 8% in the operational HWRF model at 96 hours and 14% at 72 hours. The 96-hr lead time fell within the window in which models most consistently demonstrated greater impacts, but the values do not necessarily reflect the maximum observed forecast benefit.

Table ES-1. Global Hawk dropsonde impact on 96-hour tropical cyclone track forecasts (i.e., percent improvement) for multiple storms in 2014-2016 and Hurricane Matthew in 2016, both with and without a gap in satellite coverage.

Model and Version		All Observations		Satellite Gap	
		Multi Storm (%)	Matthew 2016 (%)	Multi Storm (%)	Matthew 2016 (%)
HWRF	V2015	–	–	5%	30%
	V2017	9%	–	–	–
	HEDAS	10%	–	–	–
GFS	V2015	–	8%	–	8%
	V2017	12%	28%	–	–

The 2015 GFS version was also used by GSD to explore the impact of GH observations on midlatitude winter storms using data from a Pacific flight that sampled a storm in advance of its impacts on southern Alaska. The results again demonstrated positive forecast impacts in multiple forecast fields in the Alaskan region of impact both with and without a gap in satellite coverage, but the gains were smaller than for the tropical cyclones. The benefits approached 6% in a satellite gap and about 4% with the conventional observing system. Given the limited sample size, however, these results were not found to be statistically significant. Global scale impacts were generally neutral for all the investigations with the 2015 GFS version.

OSSE studies were conducted for both tropical and midlatitude winter storm systems. The results were broadly consistent with the OSEs in reflecting potential forecast benefits from addition of idealized GH-type observations. The overarching goal for the OSSE studies was to assess optimal sampling strategies and the relative impact of different measurement types. The hurricane OSSEs, utilizing HEDAS, suggested forecast impact could be increased by enhancing observations closer to the storm center in comparison with the broader environment. The winter OSSEs, employing GFS, suggested preferential value in wind and temperature profile observations and supported the value of focusing sampling on a

combination of computed regions of enhanced model forecast sensitivity and key meteorological features. While providing some information on the value of measurements of different environmental parameters, the studies provided limited guidance on the selection of specific sensors.

Overall, through SHOUT and the collaborative work at NCEP/EMC, several diverse, but complementary, studies consistently demonstrated significant positive forecast benefits from including targeted observations from an unmanned aircraft like the GH during high- impact weather events. The results obtained at EMC with the current operational modeling system, in particular, are highly positive and argue strongly for the potential merit of the observations. Notable forecast improvements are also observed when existing elements of our satellite observing system are withheld, simulating the value of the GH data in the event of a possible gap in polar satellite observations. The observed benefits span both regional and global models. Larger sample sizes would be desirable to increase statistical significance, but the results are highly encouraging and support the potential for forecast benefit from operational utilization of a platform like the GH. While an ultimate decision on utilization of the GH will incorporate budgetary considerations, the scientific value of the observations appears to be broadly supported.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Abbreviation	Description
3DVAR	Three-dimensional Variational Assimilation
ACARS	Aircraft Communications Addressing and Reporting System
AIRS	Atmospheric Infrared Sounder
AMV	Atmospheric Motion Vector
AOML	Atlantic Oceanographic and Meteorological Laboratory
ATMS	Advanced Technology Microwave Sounder
AVAPS	Airborne Vertical Atmospheric Profiling System
CIRA	Cooperative Institute for Research in the Atmosphere
COAMPS-TC	Coupled Ocean/Atmosphere Mesoscale Prediction System for Tropical Cyclones
CONUS	Continental United States
CrIS	Cross-track Infrared Sounder
CTL	Climate Time Line -or- Control
DA	Data Assimilation
DRA	Disaster Relief Appropriations
DROPS	Dropsonde
ECMWF	European Centre for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
EnKF	Ensemble Kalman Filter

Abbreviation	Description
ENRR	El Niño Rapid Response
EnSRF	Ensemble Square Root Kalman Filter
ESRL	Earth System Research Laboratory
ETS	Ensemble Transform Sensitivity
FSP	Frequency of Superior Performance
GDAS	GSI Hybrid Variational Ensemble Kalman Filter data assimilation system
GFS	Global Forecast System
GH	Global Hawk
GOES	Geostationary Operational Environmental Satellite
GOSA	Global Observing Systems Analysis
GPS	Global Positioning System
GRIP	Genesis and Rapid Intensification Processes
GSD	Global Systems Division
GSI	Gridpoint Statistical Interpolation
GTS	Global Telecommunication System
H3	Hurricane and Severe Storm Sentinel
HALE	High-Altitude, Long-Endurance
HAMSR	High-Altitude Monolithic Microwave Integrated Circuit (MMIC) Sounding Radiometer
HEDAS	Hurricane Ensemble Data Assimilation
HIRAD	Hurricane Imaging Radiometer
HIWRAP	High-altitude Imaging Wind and Rain Airborne Profiler
HRD	Hurricane Research Division
HWRF	Hurricane Weather Research and Forecasting
JPL	Jet Propulsion Laboratory
JPSS	Joint Polar Satellite System
MSLP	Minimum Sea Level Pressure
NCEP	National Centers for Environmental Prediction
NHC	National Hurricane Center
NMM	Nonhydrostatic Mesoscale Model
NOAA	National Oceanic and Atmospheric Administration
NPP	National Polar-orbiting Partnership
NWP	Numerical Weather Prediction
OSE	Observing System Experiments
OSSE	Observing System Simulation Experiments
PDF	Probability Density Function
RMSE	Root Mean Squared Error
RMW	Radius of Maximum Winds
RO	Radio Occultation
SFMR	Stepped Frequency Microwave Radiometer
S-HIS	Scanning High-resolution Interferometer Sounder
SHOUT	Sensing Hazards with Operational Unmanned Technology
SLP	Sea Level Pressure

Abbreviation	Description
SS	Steady State
TCI	Tropical Cyclone Intensity
TCVitals	Tropical Cyclone Vitals Database
TDR	Tail Doppler Radar
UAS	Unmanned Aircraft Systems
UTC	Coordinated Universal Time
VAD	Velocity-Azimuth Display
Vr	Doppler Velocity
VWP	Velocity-Azimuth Display Wind Profiles
WRF	Weather Research and Forecasting

ABSTRACT

The Sensing Hazards with Operational Unmanned Technology Project supported studies to evaluate the potential utility of observations from unmanned aircraft systems like the Global Hawk (GH) aircraft to improve forecasts of high-impact weather events or mitigate potential degradation of forecasts in the event of a future gap in satellite coverage. The impact assessments incorporated both data denial studies using actual GH data, termed observing system experiments, and observing system simulation experiments based on simulated observations taken from a "nature" run assumed to represent reality. The studies evaluated the effects of GH-type observations on forecasts of both tropical cyclones and midlatitude winter storms utilizing the Global Forecast System model, and regional-scale tropical cyclone forecasts based on the Hurricane Weather Research and Forecasting model. Highlights from the diverse, but complementary, studies are summarized in this document. Overall, the results consistently demonstrated positive forecast benefits from targeted observations from the GH aircraft. Addition of dropsonde observations to the current operational modeling system resulted in improvements in excess of 10 percent in tropical cyclone forecasts at longer lead times. Notable forecast improvements were also observed when existing elements of the satellite observing system are withheld, simulating the value of GH-type data in the event of a possible gap in polar satellite observations. Larger sample sizes would be desirable to increase statistical significance and any ultimate decision on utilization of the GH will incorporate budgetary considerations, but the scientific value of the observations appears to be broadly supported.

1 INTRODUCTION

Accurate forecasting of high-impact weather events to help protect lives and properties is one of NOAA's fundamental missions. Current numerical weather prediction (NWP) systems assimilate environmental observations from a wide range of observing systems to help improve forecast accuracy. Environmental satellites comprise a critical backbone of this observing system and, each year, many dedicated missions using manned aircraft are conducted to collect detailed supplemental observations of hurricanes and tropical cyclones. Unmanned aircraft systems (UAS) now offer an important new capability with the potential to further complement essential observations of high-impact weather events.

Because environmental data from satellites have become so critically important to the delivery of accurate weather forecasts for the nation, any gap in the environmental satellite system creates vulnerability for the nation's weather services. To offset this threat, the U.S. Congress included $111 million in the Disaster Relief Appropriations (DRA) Act of 2013 to test and evaluate options to mitigate the risk of potential polar-orbiting environmental satellite observing gaps. The Sensing Hazards with Operational Unmanned Technology (SHOUT) Project, led by the NOAA UAS Program, was one option funded under the DRA with the overarching goals of quantifying the influence of observations from high-altitude, long-endurance (HALE) UAS such as the Global Hawk (GH) aircraft on high impact weather prediction, and assessing the operational effectiveness of UAS to help mitigate the risk of satellite observing gaps.

To support these goals, SHOUT has two specific objectives:

> **Objective 1 -** Quantify the significance of unmanned observations to high impact weather prediction through data impact studies using Observing System Experiments (OSE) based on unmanned observations collected during prototype operational field missions and Observing System Simulation Experiments (OSSE) based on expected unmanned observing capabilities.
>
> **Objective 2 -** Quantify the cost and operational benefit of unmanned observing technology for high impact weather prediction through detailed analysis of life-cycle operational costs and constraints versus scientific benefit.

Supported activities included dedicated field campaigns, diverse data impact assessment studies by different analysis teams, and development of detailed cost and operational effectiveness analyses. This report directly addresses the first of these objectives, assessing the scientific utility of observations from a GH-type aircraft. In fulfillment of Objective 2, a separate companion cost study (Kenul et al. 2018) presents a detailed analysis of the costs, staffing, and logistical requirements associated with operational utilization of the GH in support of NOAA goals.

The high-impact weather events sampled and analyzed included tropical cyclones and

landfalling Pacific winter storms. Hurricanes and tropical cyclones are among the most potentially destructive high-impact weather events and pose a major forecasting challenge to NOAA. Major winter storms over the Pacific Ocean, including atmospheric river events, which make landfall and bring strong winds and extreme precipitation to the West Coast and Alaska are also important to forecast accurately because of their societal impact in those parts of the country. The SHOUT project supported three campaigns using the NASA GH aircraft. Deployments targeting hurricanes and tropical cyclones were conducted in both 2015 and 2016. The third campaign explored the impact of winter storms during February of 2016 in partnership with the broader NOAA El Niño Rapid Response (ENRR) Experiment (Dole et al. 2017). Detailed descriptions of these campaigns documenting the observations collected in support of the data impact studies are contained in a companion campaign summary document (Dunion et al. 2018).

The NASA GH was the UAS utilized in all campaigns for multiple reasons including capability, technical maturity, availability of previously integrated and proven sensors of relevance, and potential availability for future NOAA operational use. Further detail motivating focus of the analyses on the GH is provided in the cost study (Kenul et al. 2018). In addition to the SHOUT supported deployments in 2015 and 2016, the impact studies also employed data collected with the GH during the NASA-led Hurricane and Severe Storm Sentinel (HS3, Braun et al. 2016) investigation in 2012-2014 and the NASA Genesis and Rapid Intensification Processes (GRIP, Braun et al. 2013) experiment in 2010.

The primary sensors selected and flown during SHOUT included the Airborne Vertical Atmospheric Profiling System (AVAPS or dropsondes), High-Altitude Monolithic Microwave Integrated Circuit (MMIC) Sounding Radiometer (HAMSR), and High-altitude Imaging Wind and Rain Airborne Profiler (HIWRAP). These payloads were prioritized based both on their perceived potential to support forecast improvements and their maturity. Additional candidate sensors evaluated based on deployments in other projects included the Scanning High-resolution Interferometer Sounder (S-HIS) and Hurricane Imaging Radiometer (HIRAD). More details on the sensor characteristics and individual observations are presented in the campaign summary document (Dunion et al. 2018).

The SHOUT project directly funded impact studies at both the NOAA Earth System Research Laboratory, Global Systems Division (NOAA/ESRL/GSD) and the NOAA Atlantic Oceanographic and Meteorological Laboratory, Hurricane Research Division (NOAA/AOML/HRD). The work at AOML/HRD focused on regional modeling of hurricanes and tropical storms while the research at ESRL/GSD centered on impacts within global scale models. Additional highly valuable studies were conducted at the Environmental Modeling Center within NOAA's National Centers for Environmental Prediction (NCEP/EMC) using fully operational regional and global models under their own support in collaboration with the SHOUT project.

This document is the final report summarizing the results of the multiple studies evaluating

the potential forecast impact of GH observations conducted under or in collaboration with the SHOUT project. Preliminary SHOUT impact studies were prepared in 2014-2016 based on analyses completed at the time. This final report supersedes those interim reports, integrating relevant material with the latest results in a complete single document.

The diverse but complementary studies are found to consistently demonstrate significant positive forecast benefits from including targeted observations from a UAS like the GH during high- impact weather events. Consistent with SHOUT Objective 1, the studies are comprised of both OSEs employing data denial studies with actual observations, and OSSEs based on simulated observations from a detailed model "nature run" assumed to accurately reflect reality. The results from the data denial studies are presented in Section 2, including regional and global modeling of tropical cyclones and global modeling of winter-season storms over the Pacific Ocean. Complementary OSSE results focused on evaluating different observation types and sampling strategies for both hurricanes and winter storms are described in Section 3. The final section provides a concluding synthesis and assessment of the potential ability of UAS-based observations to enable significant positive benefits to the forecasting of high-impact weather events. The presentation is intended to help facilitate NOAA management decisions on whether to pursue future operational usage of the GH or a comparable UAS.

2 OBSERVING SYSTEM EXPERIMENT RESULTS

Observing System Experiments (OSEs) are studies enabling direct evaluation of the forecast impact of real observations collected from UAS or other platforms. These include formal data denial studies where the forecasts produced by multiple model and assimilation systems can be compared with and without the UAS observations and other components of the observing system such as satellite data. SHOUT supported OSEs related to both tropical cyclones and high-impact midlatitude weather events using global and regional models. The results of these studies represent the primary scientific deliverable of the project.

The studies, spanning a diverse range of modeling systems and weather targets, consistently demonstrate highly positive forecast impact from GH UAS observations when added both to the present full observing system and to a simulated gap in satellite observations. The majority of the results are significant, illustrating strong immediate potential benefit to forecast accuracy for high-impact weather events from GH observations. The presented studies emphasize current NOAA operational global and regional models and include results from fully operational model configurations at NCEP/EMC and local installations at individual laboratories that enable testing of additional factors such as a gap in the satellite observing system.

Results are presented first for regional modeling of tropical storms, and then for global modeling of both tropical storms and winter-season storms over the Pacific Ocean. The results are largely focused on studies performed at NOAA, but relevant studies by external groups are

also cited to highlight potential impacts of sensors not yet evaluated and on different modeling systems.

2.1 Regional Tropical Cyclone OSE Results

Assessment of the potential impact of GH observations on regional scale tropical cyclone forecasts was based on use of NOAA's Hurricane Weather Research and Forecasting (HWRF) model. This is the current operational tropical cyclone forecast model which employs the WRF Nonhydrostatic Mesoscale Model (WRF NMM) dynamical core with a storm-following grid nesting. The system supports multiple options for data assimilation including the NOAA Gridpoint Statistical Interpolation (GSI) and GSI-hybrid variational assimilation schemes and an Ensemble Kalman Filter with the Hurricane Ensemble Data Assimilation System (HEDAS, Aksoy et al. 2012, 2013) developed at AOML as a research tool. The regional HWRF studies supported by SHOUT were performed under the overall leadership of Altug Aksoy from AOML/HRD.

The studies examined the impact of different data types from the GH and compared the results obtained using different assimilation schemes. Since the HWRF model utilizes boundary conditions obtained from the GFS model, the testing was closely coupled with the global modeling activities described in Section 2.2. In collaboration with SHOUT, NCEP/EMC also conducted independent evaluations using the fully operational model configuration. Additional studies conducted independently from the SHOUT project also provide important insight into the potential value of observations from platforms like the GH.

2.1.1 Studies Employing Operational HWRF within SHOUT

Studies documenting the impact of GH observations on forecasts from the operational HWRF modeling and assimilation system are of particular value to SHOUT since they demonstrate the potential immediate impact on NOAA operations. Experiments with an operational version of HWRF were performed by James Taylor (from the Cooperative Institute for Research in the Atmosphere, CIRA) through a close collaboration between the analysis teams at AOML/HRD and ESRL/GSD. This work explicitly addressed the ability of the GH observations to fill a gap in satellite coverage. The studies employed the 2015 version of the operational HWRF model and the GSI assimilation system with the 3-dimensional ensemble-variational hybrid assimilation scheme (Wu et al. 2002; Kleist et al. 2009) operational through mid-2016. The HWRF model consists of an outer domain and two storm following inner nests (D02 and D03). A graphic illustrating the operational system with vortex relocation and data assimilation components is shown in Figure 2.1.

The results presented in this report are for Hurricane Matthew from 2016. Earlier testing was also conducted for Hurricane Edouard from 2014, but those analyses were performed prior to correction of a dry bias identified in the dropsonde data (Vömel et al. 2016) and are being repeated. Boundary conditions for the Matthew tests were provided from operational Global

Forecast System (GFS) model runs performed by the team at ESRL/GSD (see Section 2.2). The GFS runs were fully compatible with the HWRF experiments with respect to the inclusion/exclusion of both GH and satellite observations.

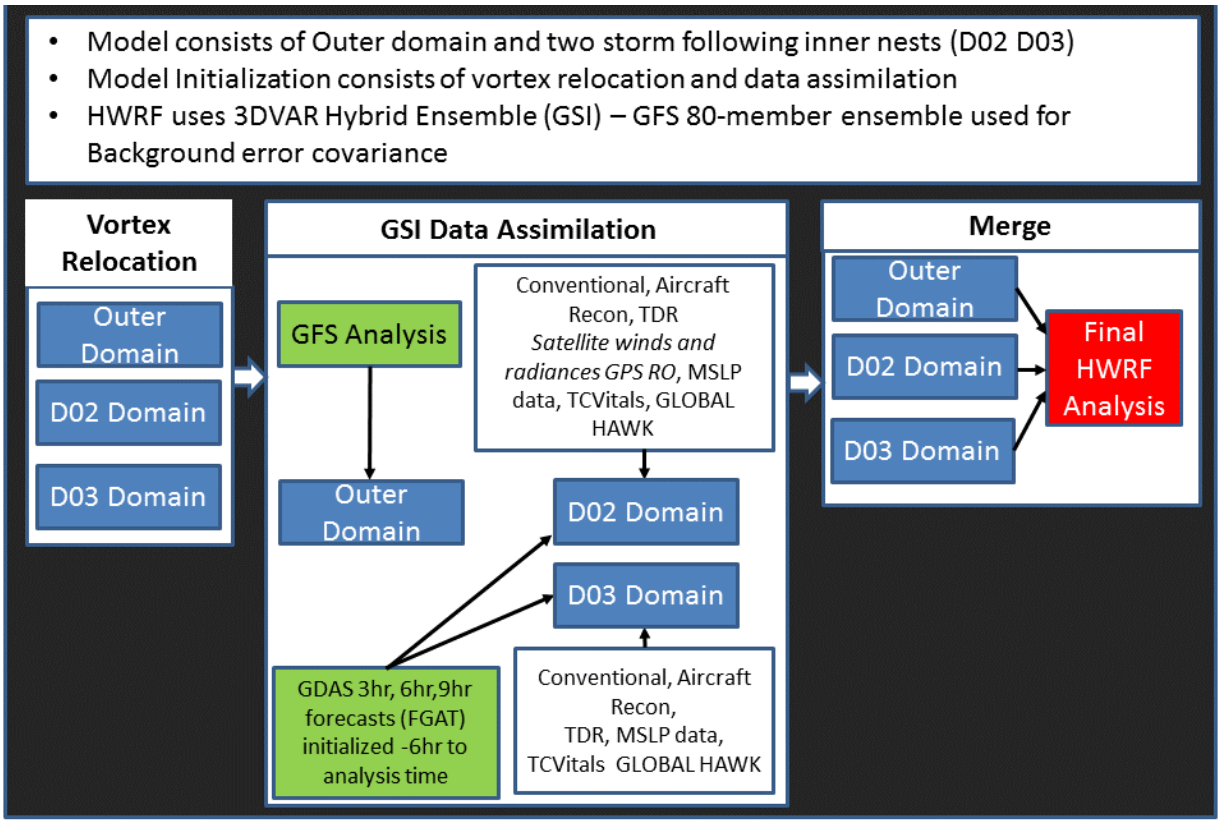

Figure 2.1. Schematic of the 2015 operational Hurricane Weather Research and Forecasting (HWRF) modeling system employed in the presented impact studies. The graphic provided by James Taylor includes the following abbreviations: Three-dimensional Variational Assimilation (3DVAR), Global Forecast System (GFS), Global Positioning System Radio Occultation (GPS RO), Gridpoint Statistical interpolation (GSI), GSI Hybrid Variational Ensemble Kalman Filter data assimilation system (GDAS), Minimum Sea Level Pressure (MSLP), Tail Doppler Radar (TDR), and Tropical Cyclone Vitals Database (TCVitals).

Several experiments were conducted to document the impact of observations from the GH both with and without elements of our current satellite observing system. The experiments examined the impact of both the GH dropsonde and HAMSR observations during cycling over the full duration of the Matthew flights, from 12 UTC on 4 October through 18 UTC on 9 October. Evaluation metrics focused on the traditional measures of forecast track and intensity but also explored the impact on the inner core structure. As with the current HWRF operational configuration, all observations within ~150 km of the storm center were not assimilated. These observations are excluded operationally because of their potential to degrade forecasts in current HWRF versions. This could limit the ultimate potential impact of the GH observations given the focus on over-storm sampling during SHOUT.

The primary study examined the potential impact of GH dropsonde and preliminary HAMSR

observations on forecasts in the event of a gap in polar-orbiting satellite coverage. A potential gap in the current satellite observing system was simulated by withholding sounding data from the Advanced Technology Microwave Sounder (ATMS) and Cross-track Infrared Sounder (CrIS) instruments onboard the Suomi NPP satellite. This reflected the primary potential satellite gap of concern which would have resulted if the Suomi NPP satellite failed before the launch of the first Joint Polar Satellite System (JPSS) satellite. The launch occurred successfully in November 2017. The ATMS and CrIS instruments are the primary sources of data from the Suomi NPP and JPSS-1 satellites used to support operational weather forecasting.

For these tests, the dropsondes were assigned a reduced vertical resolution comparable to the typical operational Global Telecommunication System (GTS) TEMP DROP messages. Unlike the traditional TEMP DROP messages which incorporate observations at specified mandatory and significant levels, the dropsonde data were averaged into 38 regularly spaced vertical levels corresponding to 25 mb spacing. The data included a measured geographic position at each level which also differs from the TEMP DROP messages where the position is provided only for the launch location. Inclusion of the position represents an important increase in information content in these data relative to operational observations.

The HAMSR retrievals of temperature and humidity were quality screened by the instrument team at the NASA Jet Propulsion Laboratory (JPL) to remove any observations that were degraded by precipitation, and averaged over 5 km x 5 km regions to facilitate data assimilation. Additionally, the data were thinned temporally to take no more than one profile every one minute. The retrievals were again interpreted on 38 regularly spaced vertical levels as for the dropsondes. Within the data assimilation scheme the retrievals were assumed to have the same error characteristics as dropsondes. This is not entirely appropriate, but was a useful first approximation in the absence of more detailed analysis. Additional work with the HAMSR data is planned in the future.

The results obtained in forecast cycles with available data are very encouraging. The impact of the GH observations on the Matthew track and intensity forecasts averaged over the three cycles (12 and 18 UTC on 5 Oct and 12 UTC on 7 Oct) when the majority of the observations were assimilated are shown in Figure 2.2. The results are presented independently for assimilation of the dropsonde and HAMSR data. The corresponding locations of the assimilated observations are shown in Figure 2.3, relative to the inner and outer model domains. Inclusion of the dropsondes results in a clear and significant reduction of the track error at lead times greater than about 36 hours. The improvement is consistently on the order of 30%. The impacts on intensity are more mixed with periods of both improvement and degradation. Inclusion of the HAMSR data also led to notable improvements in the track forecast. Improvement on the order of 20% is observed in the first 18 hours and then consistent improvement at longer lead times is observed after 60 hours. The amount of improvement at longer lead times is slightly more variable than for the dropsondes, peaking at 72 hours and then somewhat reduced toward the end of the forecast interval. The HAMSR

data did not benefit the intensity forecast, with a notable skill degradation between 18-60 hours. During this period, however, the original error in the minimum pressure forecast was already quite small.

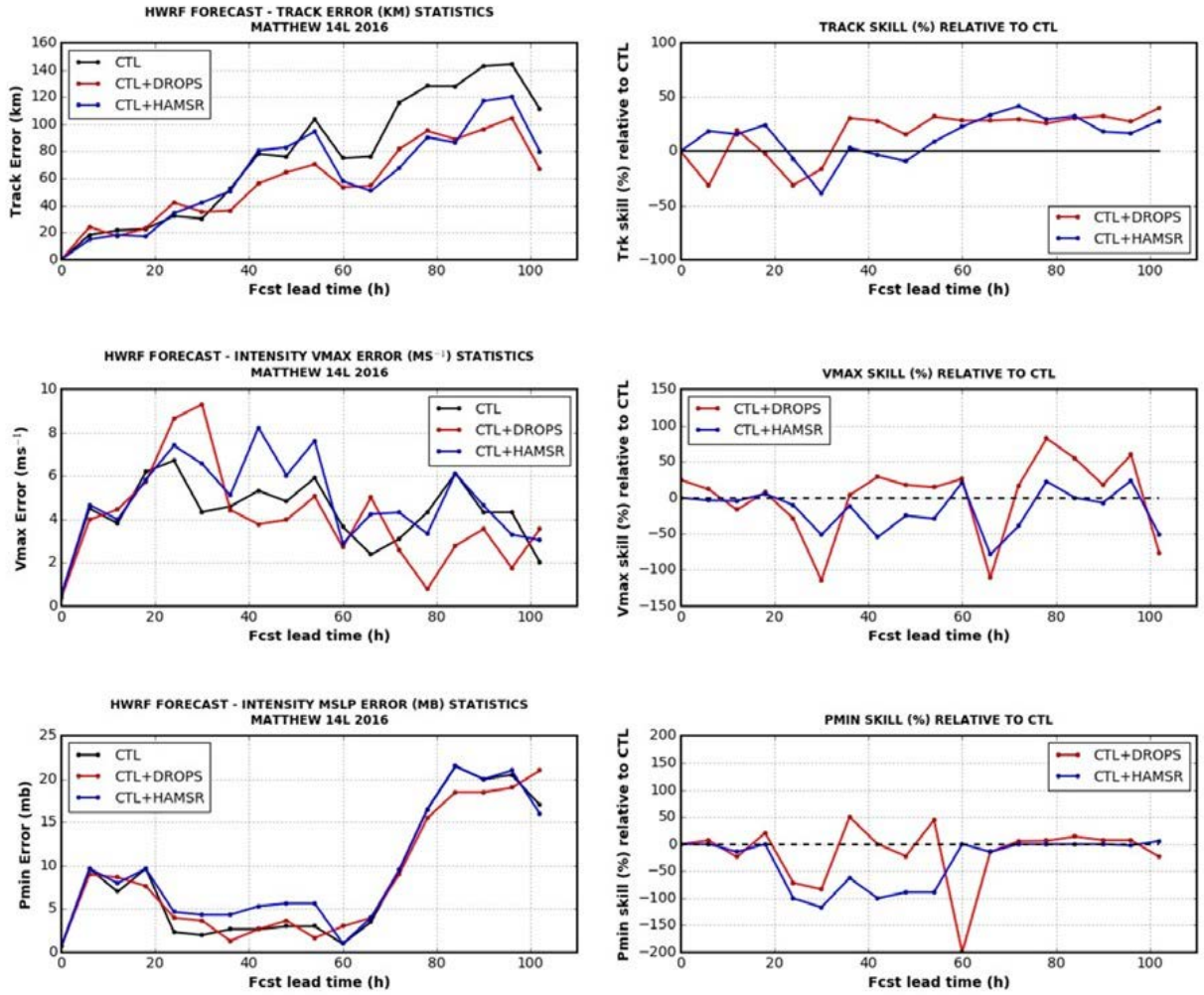

Figure 2.2. Summary of GH dropsonde impact on the operational HWRF model forecast of track and intensity for Hurricane Matthew when satellite observations were withheld. The results shown are averaged over three cycles during which dropsondes (DROPS) and HAMSR were assimilated. The left column shows the track and intensity forecasts with and without GH observations while the right column shows the corresponding percent change in forecast accuracy relative to the control (CTL) case with no GH observations. The red traces illustrate the dropsonde impact and blue traces correspond to the addition of HAMSR. Results provided by James Taylor.

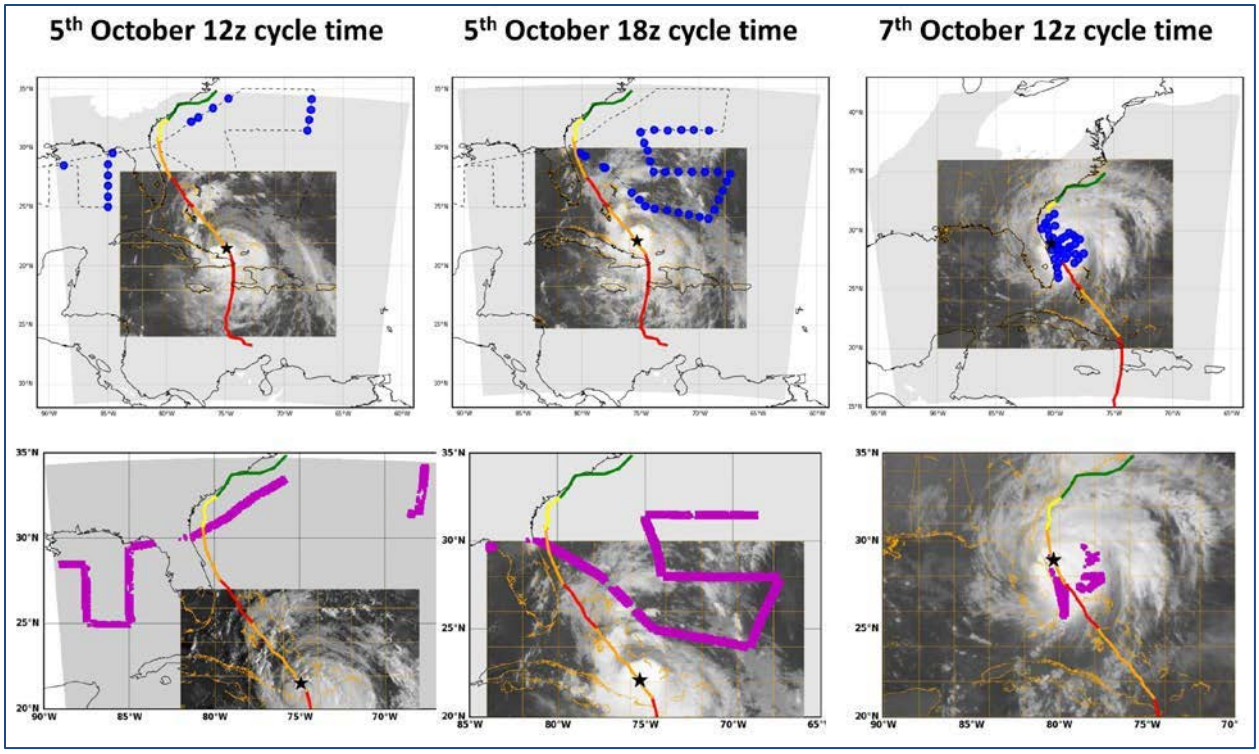

Figure 2.3. Illustration of the location of the GH observations assimilated for Hurricane Matthew yielding the impact results presented in Figure 2.2. Dropsonde locations are indicated with blue symbols in the upper panels and HAMSR retrievals by magenta symbols in the lower panels. The columns show the different model cycles. The inset image and shading show the two inner domains of HWRF. Graphics provided by James Taylor.

The improvement in the track forecast had a significant impact on the corresponding forecast precipitation along the southeast coast of the United States. The impact on the track and precipitation was examined in detail for the forecast initialized at 18 UTC on 5 October and the results are shown in Figure 2.4. Assimilation of the GH observations is observed to move the forecast track closer to the coast, more consistent with the analyzed track generated by the National Hurricane Center (NHC), particularly for the dropsonde data. With this shift, the maximum precipitation along the coast of the Carolinas moves onshore in better agreement with the observed precipitation. This represents a significant improvement in a quantity with tremendous societal impact.

The corresponding impact of the GH dropsonde observations averaged over all Matthew model cycles, including those with no observations, is shown in Figure 2.5. While generally reduced slightly, the impact of the observations on the track forecast remains very positive with a consistent benefit of near 20% at lead times beyond 30 hours. The impact on intensity remains mixed. The observations show a largely neutral impact with respect to the maximum wind speed, but some hint of potential benefit to the forecast minimum pressure between 36 and 66 hours. The results suggest the observations do provide residual value in later model cycles.

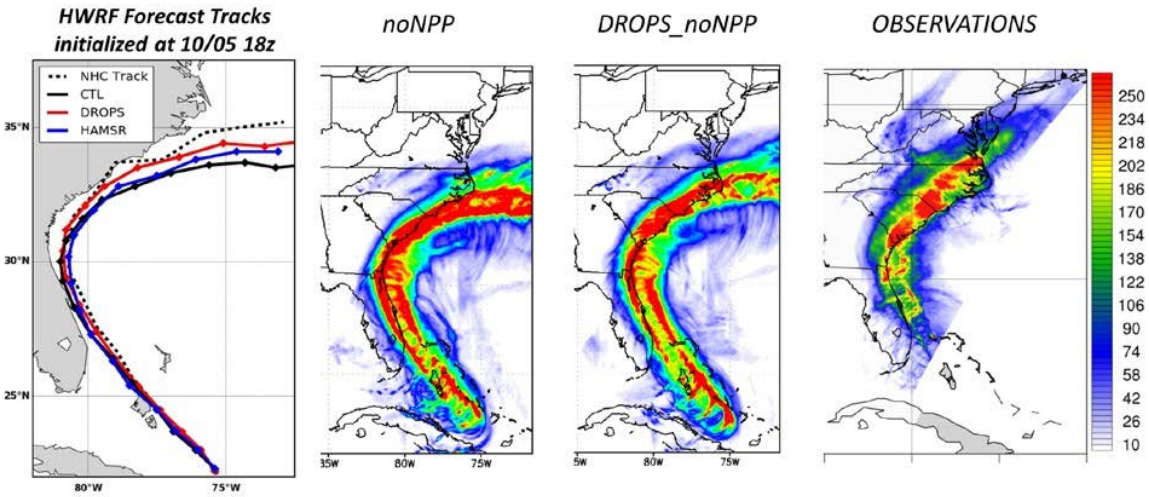

Figure 2.4. Illustration of the impact of GH dropsondes (DROPS) on the forecasted track and precipitation of Hurricane Matthew. The observed (NHC) and forecasted tracks are shown in the left most panel relative to the control (CTL) forecast with no GH observations. The second and third panels show the forecasted precipitation with and without assimilation of the dropsondes for comparison with the observed precipitation as shown in the final panel. The observations were taken from Stage IV precipitation data (Lin 2011). Graphics provided by James Taylor.

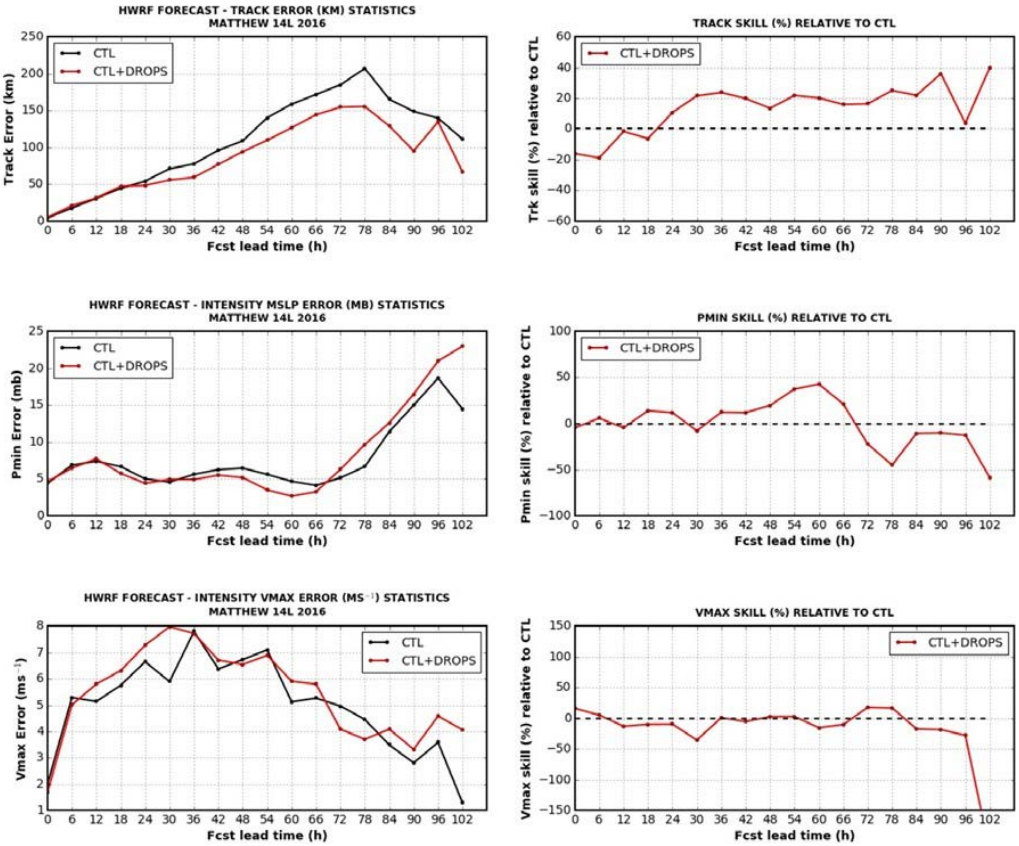

Figure 2.5. As in Figure 2.2, but for the impact of dropsonde observations when averaged over all forecast cycles during Hurricane Matthew. Results provided by James Taylor.

Overall, these results are highly positive with respect to the potential value of GH observations to operational HWRF forecasts in the absence of regular satellite observations, particularly during those periods when the observations are directly assimilated. The large improvements in the track forecast from assimilation of both the dropsondes and preliminary HAMSR data are particularly notable. Further work with the HAMSR data on thinning procedures and error characteristics is underway to better understand its potential impact, particularly on intensity. While the results are for a single storm only, complementary work at NCEP/EMC described in the following section assesses the dropsonde impact on a much larger sample size in the presence of the current full observing system.

2.1.2 Studies Employing Operational HWRF at NCEP/EMC

Work ongoing at NCEP/EMC extended these results, looking more broadly at the overall impact of GH dropsonde observations on the most recent H217 version of HWRF with the latest data assimilation system which was implemented in 2017. These activities were closely coordinated with the NCEP/EMC investigation of the impact on the operational GFS model described in section 2.2.2. The operational GFS boundary conditions utilized within the HWRF modeling system were entirely consistent with respect to inclusion or exclusion of the GH dropsonde data. The results provide the most direct indication of the potential value of the GH data if implemented into the current operational forecast system along with all standard existing observations. The work was conducted by Jason Sippel and Kate Howard under the leadership of Vijay Tallapragada.

In this analysis, the impact of GH dropsonde data obtained directly from the TEMP DROP messages submitted in real time through the GTS was evaluated relative to the operational forecasts obtained assimilating all normal conventional and reconnaissance observations. In the TEMP DROP messages, dropsonde measurements were reported only at specified mandatory and significant levels and not at the full measurement rate. With this approach, the data also contained only the launch location and lacked any information on absolute position during the descent. Observations within ~150 km of the storm center are still excluded in this version of HWRF. The overall investigation with the GFS evaluated the impact of all GH dropsondes deployed during the HS3 campaign in 2014 and the SHOUT hurricane campaign in 2016. The HWRF analysis is still ongoing and the results available thus far are from the 2016 flights up through part of the sampling of Hurricane Matthew.

Previous results shared by Jason Sippel examining the impact of GH observations on operational HWRF forecasts of individual storms had largely failed to show any consistent positive impact of the data, particularly for the forecast track. In those studies, however, the boundary conditions were drawn from the original operational GFS model runs in which GH dropsonde data were not assimilated. Any impact of the GH dropsondes on the broader environment and HWRF boundary conditions were not captured in those results. In the current results, the combined impact of the GH observations on the GFS and HWRF models

are fully realized.

The results summarizing the impact of the 2016 observations on forecast track and intensity skill shown in Figure 2.6 demonstrate significant improvements in both track and intensity. The impact on the forecast intensity is positive throughout the forecast period with the greatest impacts at longer lead times. The benefit peaks at around 14% at a lead time of 72 hours. For track the forecast is only improved at longer lead times with a peak improvement of around 10% at 96-hour lead. The forecast is degraded on average at the shortest lead times but Jason Sippel did not believe this to be a major concern because of issues associated with initial starts of HWRF for new storm systems. Interestingly, examination of individual forecast runs revealed that some of the forecast improvements obtained for Matthew were the result of earlier sampling during the Karl flights. Another specific observed benefit was a major improvement in the forecasted recurvature of Gaston following assimilation of the dropsonde observations from the first Gaston mission.

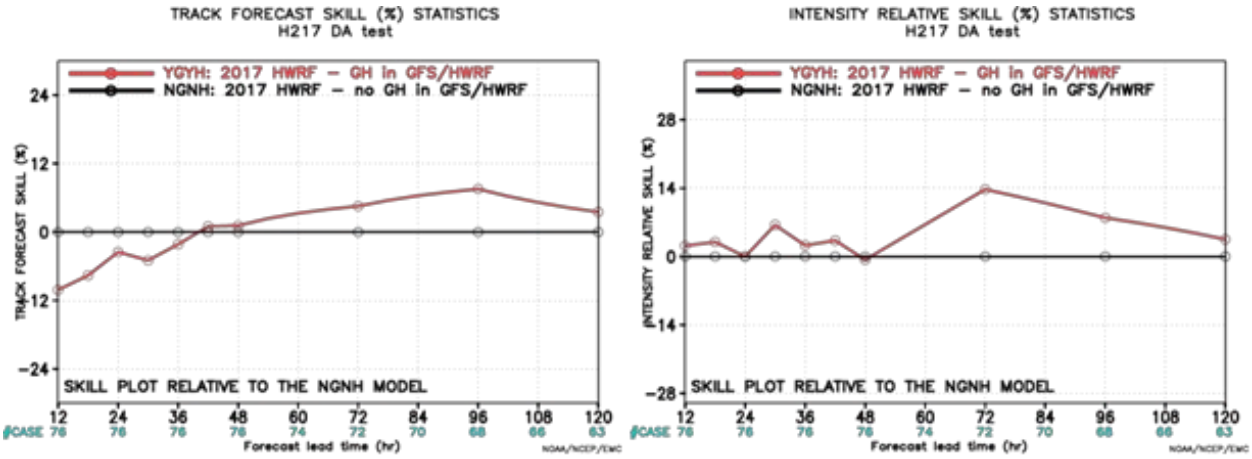

Figure 2.6. Average impact of GH dropsondes on operational H217 HWRF forecasts of track and intensity at NCEP/EMC for sampling during the 2016 hurricane missions through part of the Matthew flights. The results adding GH dropsondes (red traces, coded YGYH) represent improvements in the forecast skill for track (left) and intensity (right) relative to control runs performed assimilating all conventional observations including reconnaissance aircraft (black traces, coded NGNH). Graphics provided by Jason Sippel.

The results represent some of the most significant hurricane forecast intensity improvements observed in the SHOUT analyses. The findings are very encouraging because of their immediate potential applicability to NOAA operations and the historical difficulties in demonstrating significant improvements to intensity forecasts.

Future activities planned at NCEP/EMC that could yield even further improvements include evaluating the impact of incorporating dropsonde position information during decent, incorporating full resolution dropsonde data from BUFR rather than TEMP DROP messages, and including observations near the storm center.

2.1.3 Studies Employing HWRF and HEDAS

Studies conducted at HRD included distinct experiments focused on evaluating the impact of GH data from different sensor types on the accuracy of forecasts of hurricanes and tropical storms. Studies examined the value of profiles measured by dropsondes, retrievals of temperature and humidity from the S-HIS and HAMSR, and surface wind retrievals from the HIRAD.

The primary tropical cyclone modeling system employed at HRD utilizes the HWRF model and their own HEDAS system. The model is similar to that employed operationally at NCEP, but the data assimilation system does have key differences. The HEDAS system uses a square-root ensemble Kalman filter (e.g., Whitaker and Hamill 2002) and supports hurricane inner-core data assimilation for high-resolution vortex initialization (Aksoy et al. 2012, 2013). The function of the system is illustrated schematically in Figure 2.7. While the results obtained from these experiments do not necessarily reflect the potential impact of the GH data on the current operational modeling system, they are relevant for highlighting the possibility of obtaining further improvements through incorporating more data near the storm center and through more frequent cycling.

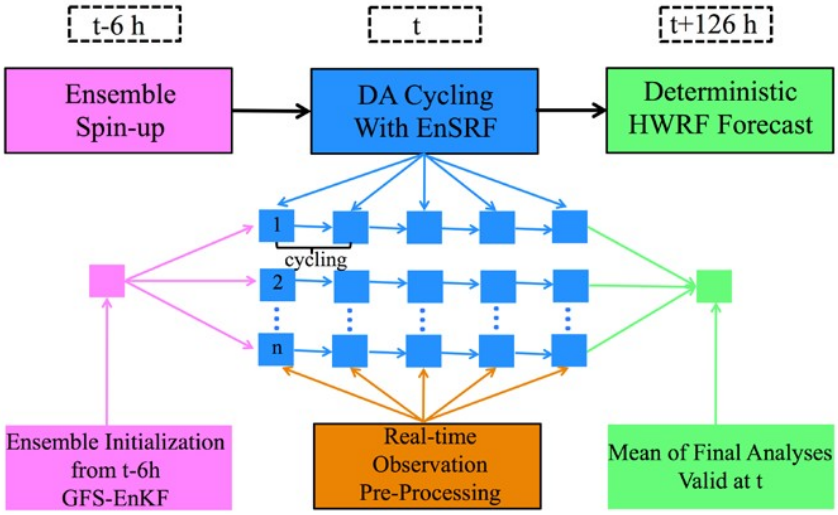

Figure 2.7. Schematic illustration of the operation of HEDAS, where DA = data assimilation, EnKF = ensemble Kalman filter, and EnSRF=ensemble square root Kalman filter. Graphic provided by Altug Aksoy.

In the HEDAS studies to date, all modifications to the assimilated data were performed within the HWRF model domain only. The HWRF model further relies on external boundary conditions provided from the GFS model. In these studies, the data assimilated within GFS was not changed and the operational GFS runs were used for the boundary conditions. As a result, the results may not reflect the full possible impact of the GH data as the observations could also affect the external boundary.

A primary study examined the composite forecast impact of GH dropsonde observations collected over tropical cyclones during both the SHOUT and HS3 campaigns. The analyses considered ten storms as shown in Figure 2.8. This work was led by Hui Christophersen of HRD. The dropsonde data was taken from the traditional TEMP DROP messages with the corresponding reduced vertical resolution (mandatory and significant levels), but additional position information at each level was estimated from the launch position and measured winds. Other observations assimilated within the HWRF domain in the control for this study included atmospheric motion vectors (AMVs, derived from Geostationary Operational Environmental Satellite [GOES] observations), GPS radio occultation profiles, and profiles of temperature and humidity retrieved from the Atmospheric Infrared Sounder (AIRS) flying on the NASA Aqua satellite.

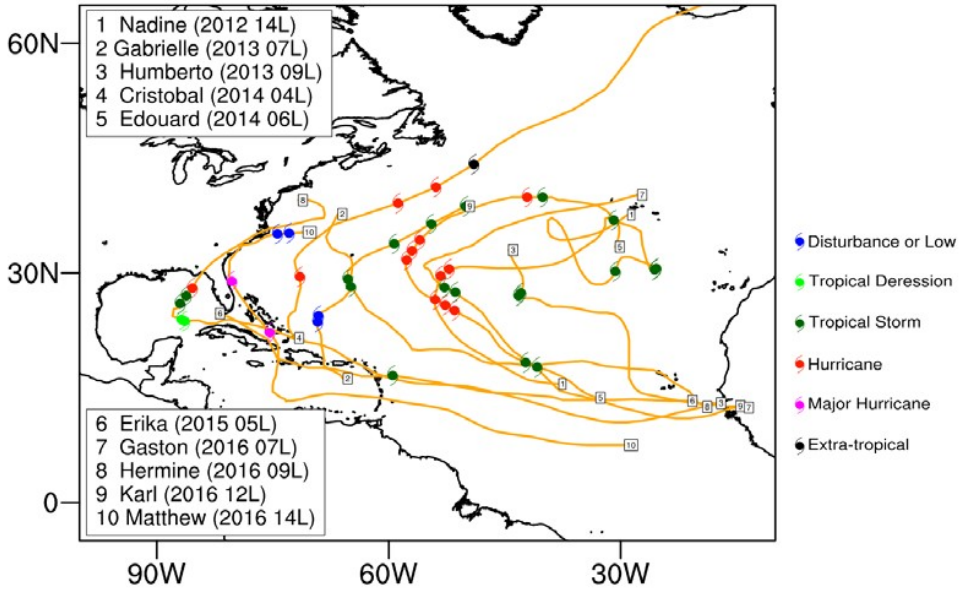

Figure 2.8. Illustration of the ten tropical cyclones and corresponding GH missions incorporated in the composite dropsonde impact assessment. The storm location and strength at the time of the GH mission are indicated by the colored symbols. Graphic provided by Hui Christophersen.

This composite study is an extension of the preliminary dropsonde impact study conducted by HRD (Christophersen et al. 2017) that demonstrated the relative value of GH dropsondes depending on the presence of other aircraft reconnaissance observations in the inner core. When other inner-core reconnaissance data were available, Christophersen et al. (2017) found that GH dropsondes in the near environment led to additional improvements in the track forecasts, whereas in the absence of other inner-core reconnaissance data, GH dropsondes in the inner core resulted in the greatest impact on intensity forecasts.

The dropsonde impact on the initial analyzed storm position and intensity was evaluated first. The results illustrating the improvement in analysis skill relative to assimilation of all other observation types is shown in Figure 2.9 stratified by storms that were either steady state (SS)

or were weakening or intensifying (non-steady state, non-SS). The determination of steady state was based on whether the intensity was observed to change by 20 knots or more over a 24-hour period. Addition of the GH dropsondes improved the initial conditions on average for all cases, but more so for the cases where the storm's intensity was changing. For the non-steady state periods, addition of the observations improved the analyzed position and intensity by more than 25%. Providing a better initial state is expected to be an important factor in improving the subsequent forecast quality.

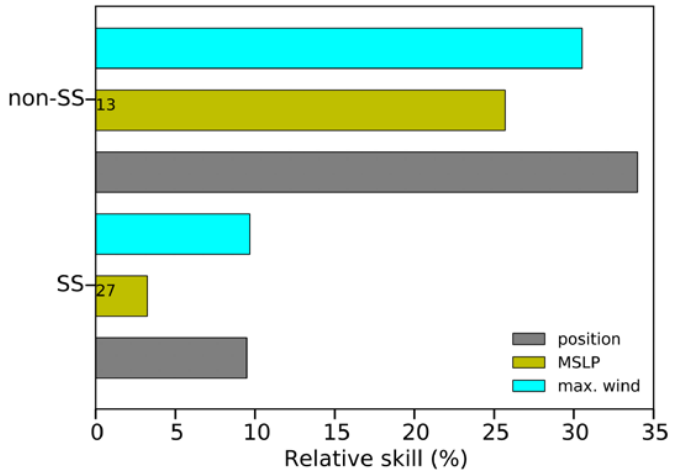

Figure 2.9. Illustration of the change in analysis skill in HWRF-HEDAS resulting from the assimilation of GH dropsondes for steady state (SS) and non-steady state (non-SS) storms drawn from the ten-storm composite investigation. Results are presented separately for the analyzed storm position, minimum sea level pressure (MSLP), and maximum wind speed. The number of cases is labeled as well. Graphic provided by Hui Christophersen.

The GH data impact on the corresponding forecasts was similarly encouraging. The average impact of the added dropsonde observations on the HEDAS forecasts was analyzed separately for cases depending on whether other reconnaissance data was collected from manned aircraft as presented in Figure 2.10. While the track forecasts were generally improved for all cases, the forecasts of intensity and minimum pressure were generally improved only where there were no other reconnaissance observations. The improvement to the track forecasts is frequently highly significant with values commonly between 10-30%. With other reconnaissance data, the improvement is greatest at longer lead times, consistent with results obtained from the operational HWRF model. With no other reconnaissance data, the observations show an increased impact at shorter leads. For intensity, the impact in the absence of other reconnaissance data is greatest at longer lead times and approaches peak values of over 30%.

The results indirectly support the value of observations closer to the storm center. For earlier tropical storms (typically from HS3), the GH observations tended to be more heavily weighted to the surrounding environment. The more recent SHOUT observations included denser

sampling near the center of weaker storms as well as hurricanes. While the forecast improvements with HEDAS were weaker on average for tropical storms than hurricanes, the track improvements for the 2016 tropical storm cases alone were quite large and significantly greater than for the earlier storms (Figure 2.11). These forecast improvements coincide with a greater reduction in the error in the initial analyzed storm position for the 2016 cases.

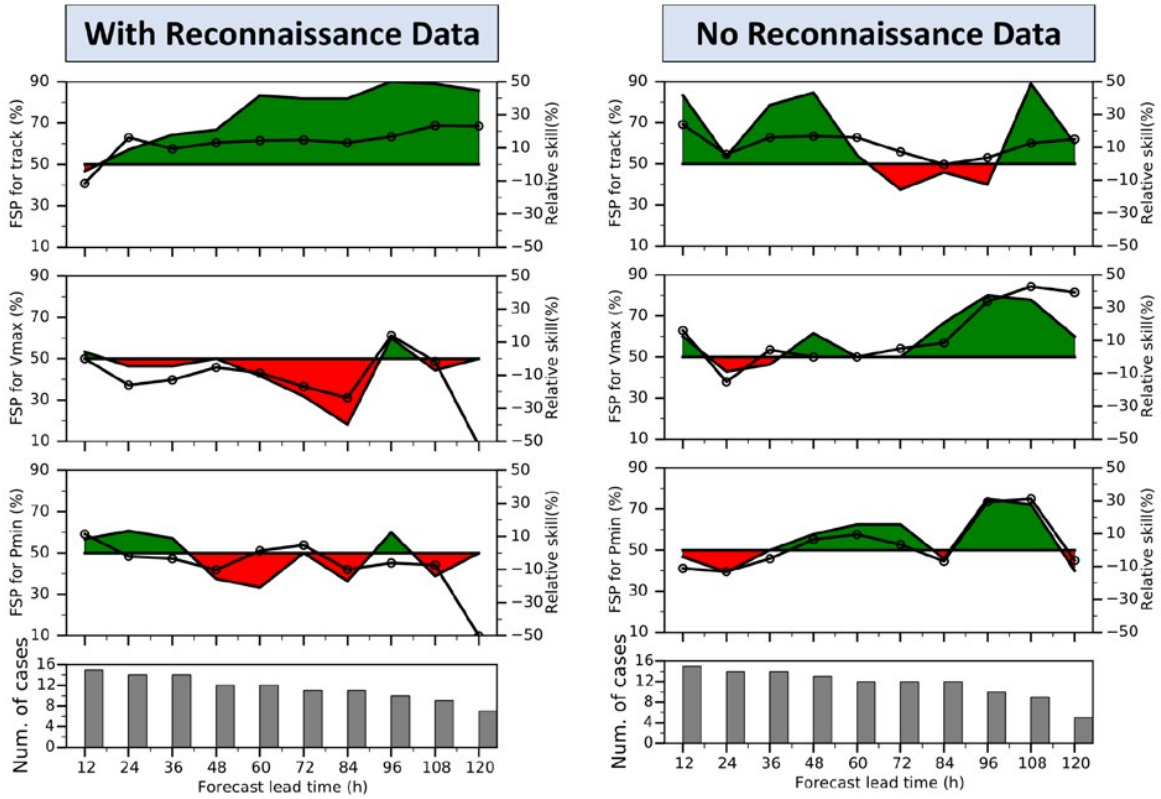

Figure 2.10. Summary of GH dropsonde impact on the HWRF-HEDAS forecasted track and intensity error for the composite analysis of ten storms. Results are stratified based on whether other aircraft reconnaissance was available (left) or not (right). The upper panels show the results for track error; the middle panels for intensity error as reflected by maximum winds; and the lower panels for minimum pressure. Results are shown both for frequency of superior performance (FSP), left axis and red/green shading, and percent improvement in forecast error, right axis and black trace with circular markers. Graphics provided by Hui Christophersen and Altug Aksoy.

To investigate the ability of GH dropsonde data to supplement or fill in for satellite data within the HWRF-HEDAS system, an additional experiment was conducted for flights from the ten-storm composite where both dropsonde measurements and AIRS retrievals of temperature and humidity were available. Comparing inclusion and exclusion of AIRS retrievals served as a proxy for estimating the value of GH observations in the absence of components of the satellite observing system. In this experiment, the relative impact of the AIRS retrievals and GH dropsonde observations were compared when added to a control experiment in which only satellite-derived AMVs were assimilated.

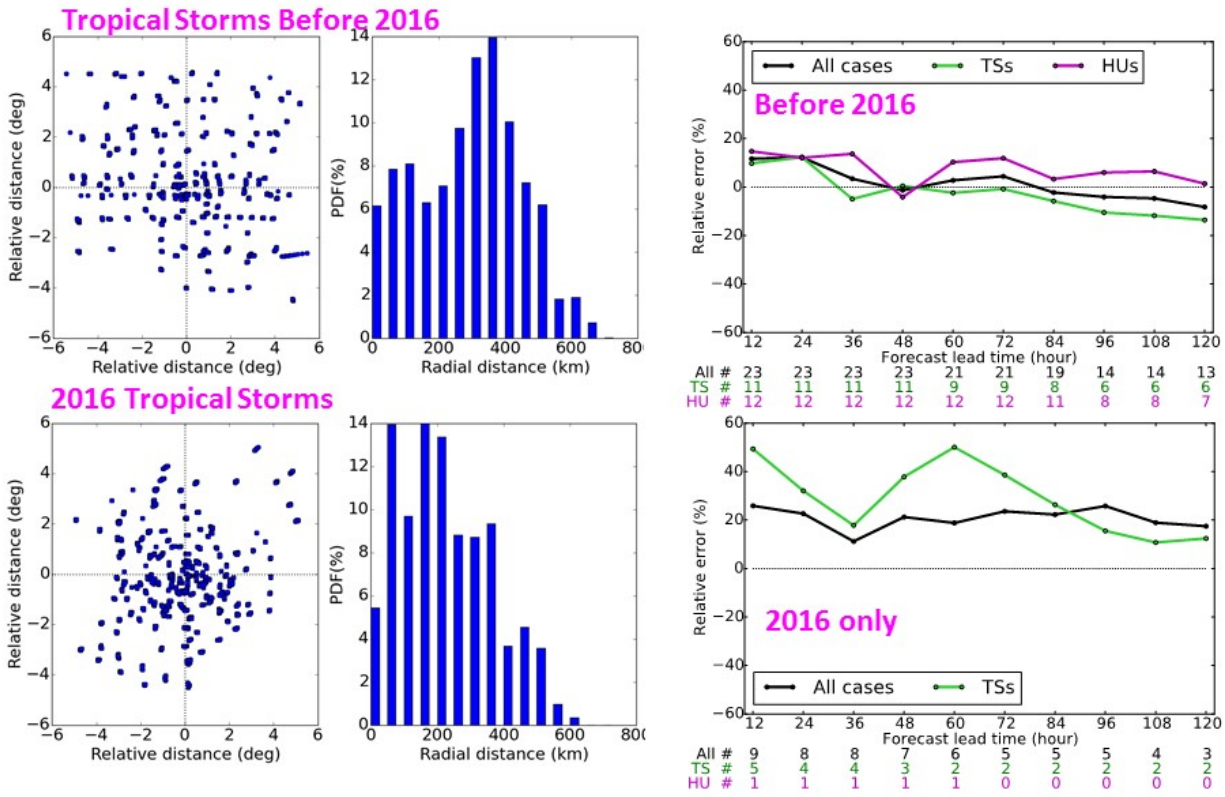

Figure 2.11. Illustration of the impact of differences in typical tropical storm (TS) sampling strategies on forecast cyclone position errors in HWRF-HEDAS. The upper row shows results for storms prior to 2016 while the lower row shows results from 2016 only. The left panels show the dropsonde positions relative to the storm center while the middle panels show probability density functions (PDFs) of the distance between the drop location and storm center. Sampling in 2016 clearly has a greater percentage of dropsondes deployed closer to the storm center. The impact of the dropsondes on position error improvement (right panels) is observed to be much more positive for the sampling in 2016. Prior to 2016, the corresponding results for hurricane (HU) strength storms is shown for comparison. Graphics provided by Hui Christophersen.

The results, shown in Figure 2.12, illustrate that GH dropsonde observations can powerfully supplement satellite observations. Addition of the dropsonde data (blue and purple traces) generally results in improved forecast skill over the control case for both track and minimum pressure. Moreover, the impact of adding of the dropsonde data alone (blue) commonly exceeds that of the AIRS data (green). Maximum positive forecast benefit is obtained through addition of both dropsondes and AIRS retrievals, but comparing the blue and purple traces illustrates that addition of the dropsonde data alone preserves most of the forecast gains. Only the track forecast at lead times in excess of 80 hours is degraded relative to the control when only dropsondes are added. A key factor in achieving the maximum benefit through assimilation of both the dropsondes and AIRS retrievals is their complementary sampling in the presence of clouds and gaps between satellite swaths.

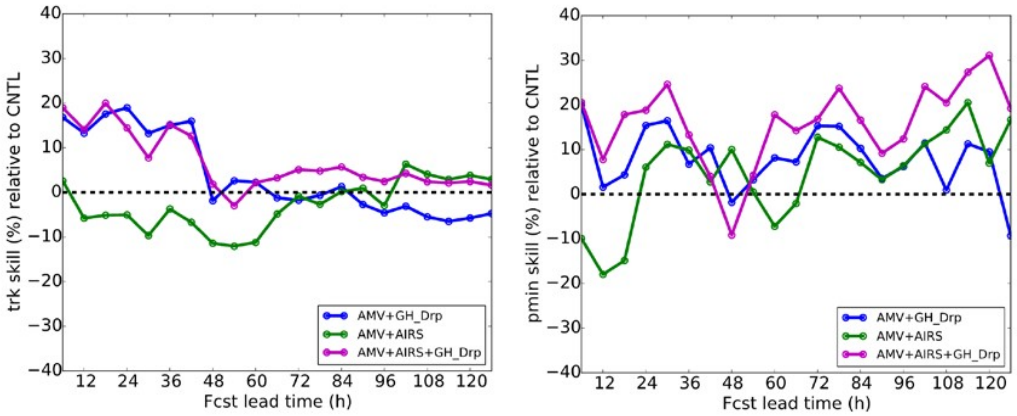

Figure 2.12. Results summarizing the impact of Global Hawk dropsondes (GH_Drp) with and without the contribution of satellite observations from the Atmospheric Infrared Sounder (AIRS). The results are averages over 19 cycles from the 10-storm composite in which both dropsonde observations and AIRS retrievals were available. The panels show the relative impact on the forecast track (left) and minimum pressure (right) skill. The colored traces show the results of the different experiments relative to a control assimilating only satellite Atmospheric Motion Vectors (AMVs) as denoted in the legend. Graphics provided by Hui Christophersen and Altug Aksoy.

An initial experiment examined the impact of HAMSR data within the HWRF-HEDAS system. Preliminary HAMSR retrievals of temperature and humidity corresponding to the Hermine 2016 flights were provided by the instrument team at JPL. As for the operational HWRF investigation in Section 2.1.1, the retrievals were screened for precipitation effects and averaged into 5x5 km bins. No further temporal thinning was required within the HEDAS system. The results, presented in Figure 2.13, suggest that the HAMSR retrievals can potentially have a positive impact on the track forecast, even exceeding that provided by the dropsondes. The impacts on intensity are mixed, but this represents a small sample size. More work is clearly required, but these initial results are encouraging.

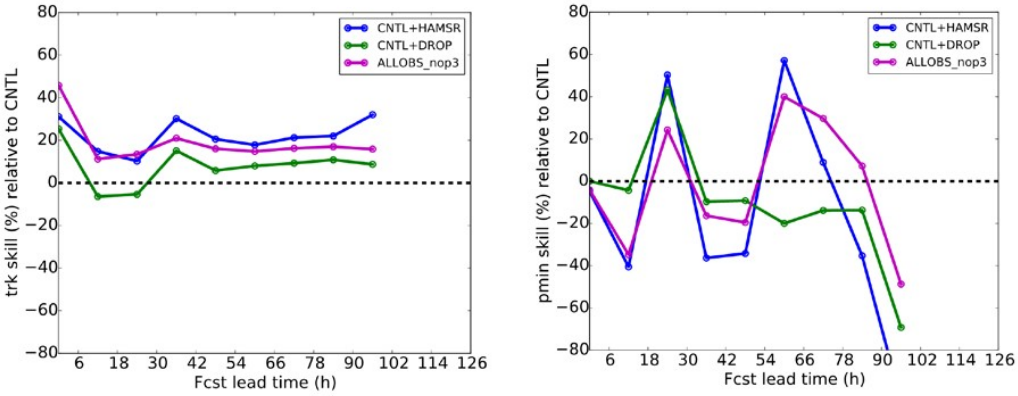

Figure 2.13. Results summarizing the impact of HAMSR retrievals within HWRF-HEDAS. The results are averages over 5 model cycles from the 2016 Hermine flights. The panels show the relative impact on the forecast track (left) and minimum pressure (right) skill for addition of the HAMSR (blue) and the dropsonde observations (DROP, green) to the control (CNTL) forecast. The magenta trace reflects addition of both HAMSR and dropsondes. Graphics provided by Hui Christophersen and Altug Aksoy.

Analysis of data collected during four Edouard missions conducted in 2014 during HS3 enabled an additional preliminary assessment of the potential impact of remotely sensed profiles from the S-HIS when flown on the GH. The S-HIS, which provides infrared-derived vertical profiles of temperature and humidity closely resembling the capabilities of CrIS, was flown on the GH during HS3 but not during SHOUT. The distribution of assimilated observation types across the different model cycles is shown in Figure 2.14. The cycles which also had WP-3D data were excluded because of the large volume of Tail Doppler Radar (TDR) data and to focus on the GH observational impact. For Edouard, addition of the S-HIS retrieved profiles to the GH dropsonde thermodynamic observations has limited impact as seen in Figure 2.15. While the overall impact of adding the GH observations is very positive, notable further improvement to forecast accuracy from addition of the S-HIS data is only observed in the forecast maximum winds at lead times less than 60 hours. There is some suggestion of a small reduction in track error beyond 60 hours but it is unclear if this is significant. These results are limited in nature and are being explored further. Analysis of one single model cycle from within the set showed the S-HIS retrievals to have a greater positive impact on the forecast accuracy at longer lead times than the dropsonde data, and that the forecast was degraded in some cases by adding the dropsonde data to the S-HIS observations. As the relative independent impact of the observations is investigated, potential issues with consistency and compatibility between the dropsonde and S-HIS observations may need to be considered.

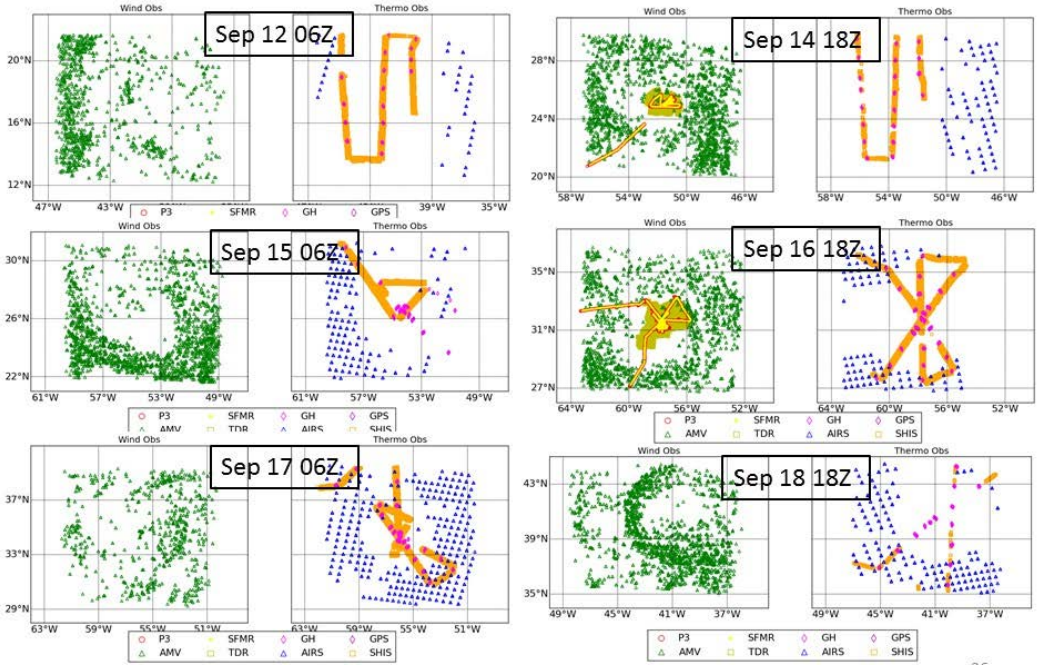

Figure 2.14. Distribution of wind (left) and thermodynamic (right) observations from different platforms and sensors for the HWRF-HEDAS model cycles evaluated during Hurricane Edouard. Observation types include AIRS retrievals (AIRS), GOES-derived atmospheric motion vectors (AMV), P-3 Stepped Frequency Microwave Radiometer (SFMR) retrievals, P-3 Tail Doppler Radar observations (TDR), GH dropsondes (GH), GPS radio occultation derived values (GPS), and GH S-HIS retrievals (SHIS) as noted by the legend at the bottom of the figure were. Graphics provided by Hui Christophersen.

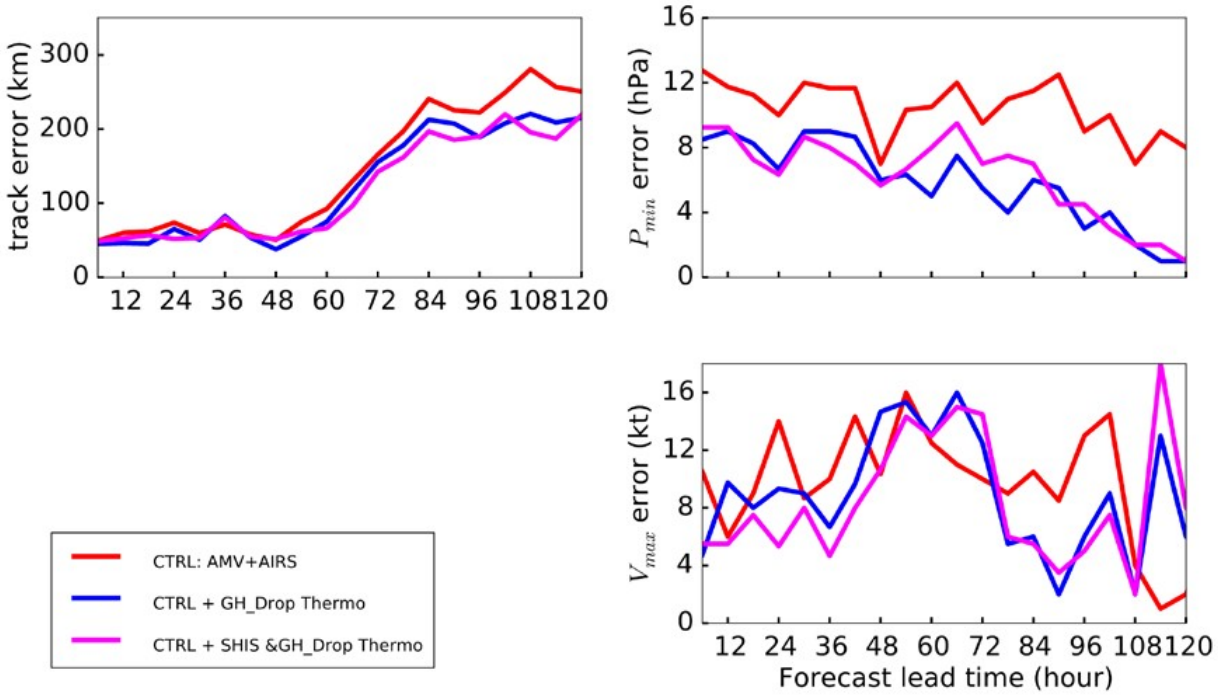

Figure 2.15. Illustration of S-HIS observation impact on HWRF-HEDAS forecasts of track and intensity for Hurricane Edouard. Results are averaged over the model cycles from Figure 2.14 without P-3 TDR data. Colored traces indicate the control (CTRL) run with AMVs and AIRS retrievals (red), the CTRL plus the Global Hawk dropsonde thermodynamic (GH_Drop Thermo, blue) observations, and the CTRL plus both GH dropsondes and S-HIS retrievals (magenta). Graphics provided by Hui Christophersen.

The potential impact of HIRAD data was assessed by Kathryn Sellwood of HRD using observations of Hurricane Joaquin in 2015 collected from three flights of the NASA WB-57 during the Tropical Cyclone Intensity (TCI) experiment. These observations are described in the operational summary (Dunion et al. 2018). HIRAD surface wind speed retrievals were provided by Dan Cecil from NASA Marshall Space Flight Center. The provided retrievals were further averaged to 5- km resolution for assimilation, which also served to reduce any residual striping in the data. The other standard observations assimilated by HWRF-HEDAS in the control study included Air Force reconnaissance aircraft flight level and dropsonde data, AMVs, Aircraft Communications Addressing and Reporting System (ACARS) data, and rawinsonde data.

Assimilation of the HIRAD retrievals within HEDAS significantly improved the initial analyzed surface structure of Hurricane Joaquin. Results summarizing the impact on the analysis are shown in Figure 2.16. The mean error in the maximum wind speed, minimum sea level pressure (MSLP), and radius of maximum winds are all significantly reduced through addition of the HIRAD data. The analyzed radius of 34 knot winds is also improved while the analyzed 50 and 64 knot wind radii are slightly degraded. Graphical comparison of the analyzed surface wind speed field with and without assimilation of HIRAD reflects the significant change to the

vortex structure, with inclusion of HIRAD resulting in a stronger, more realistic vortex.

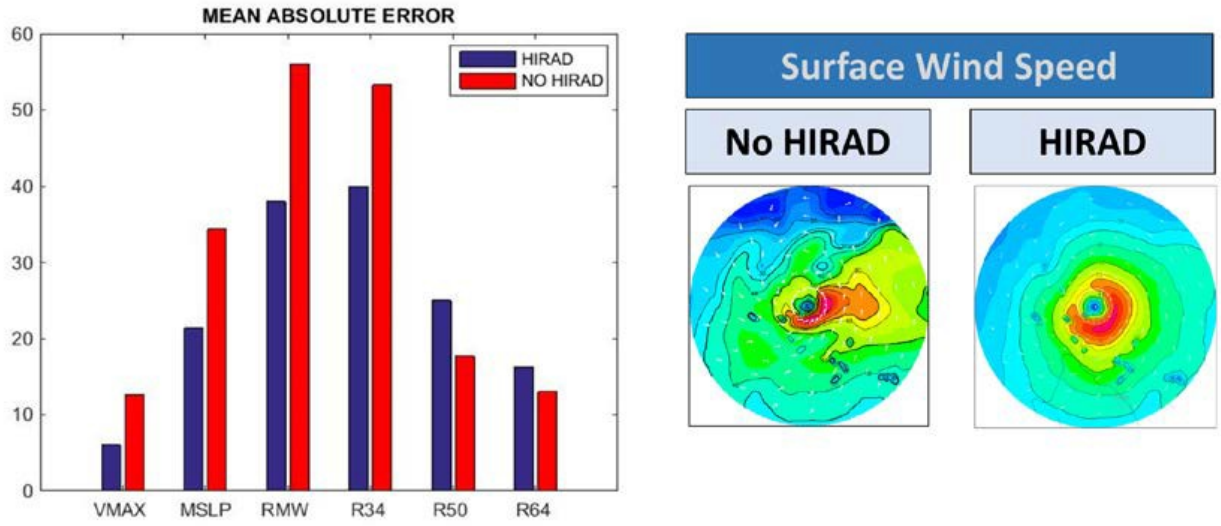

Figure 2.16. Illustration of the impact of assimilation of HIRAD surface wind retrievals on the analyzed surface structure of Hurricane Joaquin (2015). The left panel quantifies changes to the analyzed maximum wind speed (VMAX), minimum sea level pressure (MSLP), radius of maximum wind (RMW) and radii of 34- (R34), 50- (R50) and 64-knot (R64) winds. The right panel compares the spatial structure of the analyzed surface wind with (right) and without (left) assimilation of the HIRAD data. Graphics provided by Kathryn Sellwood.

The positive impacts of assimilating HIRAD data extended further into the HWRF-HEDAS forecasted track and intensity as shown in Figure 2.17. Error in the forecasted track is reduced out through 96 hours while the intensity forecasts are improved through36 hours for the maximum wind speed and through 24 hours for the MSLP.

Overall, the multiple analyses are quite positive with respect to the potential value of adding different GH observations within the HWRF-HEDAS system. GH observations, both from dropsondes and remote sensors, consistently had a significant impact on the analyzed storm structure, especially during periods of rapidly changing storm amplitude. Impacts on forecasts were smaller and more mixed, but were still largely positive, particularly for the track error. Larger sample sizes are clearly required for the remotely sensed data, but the results motivate further investigation to determine if the positive impacts are robust. The full forecast impact of the GH observations could potentially be even greater if consistently modified GFS boundary conditions are incorporated.

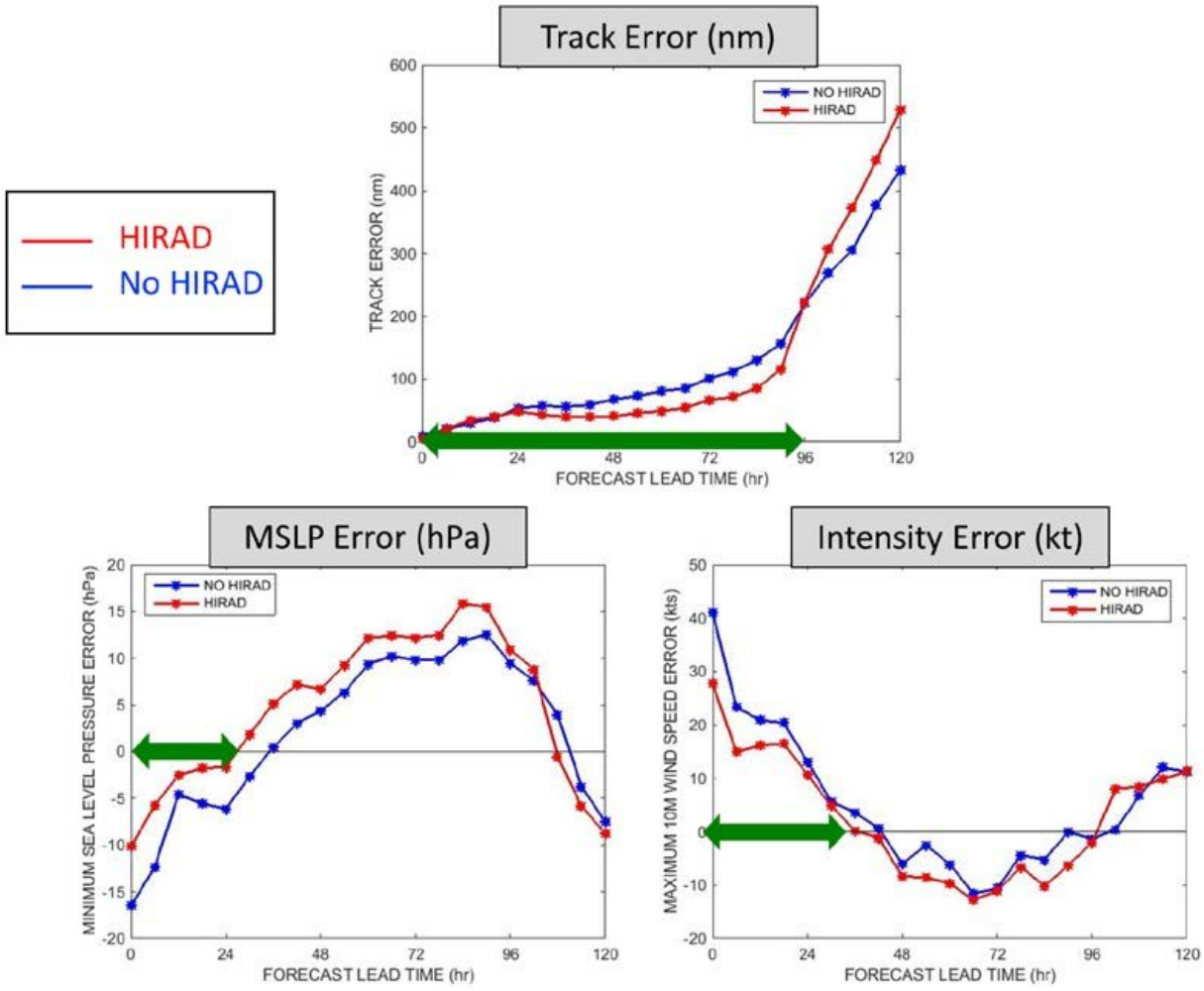

Figure 2.17. Illustration of the impact of assimilation of HIRAD surface wind retrievals on forecasts of Hurricane Joaquin (2015) within HWRF-HEDAS. The upper panel shows the impact on track errors while the bottom panels show the impact on minimum sea level pressure (MSLP) and the intensity reflected by maximum wind speed. Traces show the results with and without HIRAD as indicated in the legend. The green arrows denote the periods with forecast improvement. Graphics provided by Kathryn Sellwood.

2.1.4 Dropsonde Impact on the Navy COAMPS-TC Model

While decisions on future utilization of GH-like capabilities in NOAA will ultimately be based largely on the impact of the data within the operational GFS and HWRF modeling systems, it is instructive to also consider the observed impact on different models employed operationally by other agencies. One of the most promising initial studies on the potential positive impact of GH observations on tropical cyclone forecasts was led by Dr. James Doyle from the Naval Research Laboratory using the Navy's Coupled Ocean/Atmosphere Mesoscale Prediction System for Tropical Cyclones (COAMPS-TC). This study examined the impact of the assimilation of GH dropsonde measurements on forecasts of Hurricane Nadine in 2012. The experiments were run for the 9-day period of 19-28 September, which included 3 flights of the GH during the first field year of HS3. As demonstrated in Figure 2.18 provided by James Doyle,

assimilation of the Global Hawk dropsonde data resulted in large improvements in both the intensity and track forecast skill. The track prediction was improved by over 50 nautical miles at a lead time of 120 h. On a relative basis, the improvement in the forecasts of intensity, as reflected by the maximum wind speed and minimum sea level pressure, was even greater than those for the track. The maximum wind error was reduced from 15 knots to less than 10 knots at 48 h lead time with similar reductions at longer lead times. The corresponding error in the central pressure was reduced from 16 hPa to 10 hPa at a 48-h lead time. Since the dropsonde measurements at the lowest levels were negatively affected by telemetry interference issues during these specific flights, much of the forecast improvement was hypothesized to come from the upper level observations.

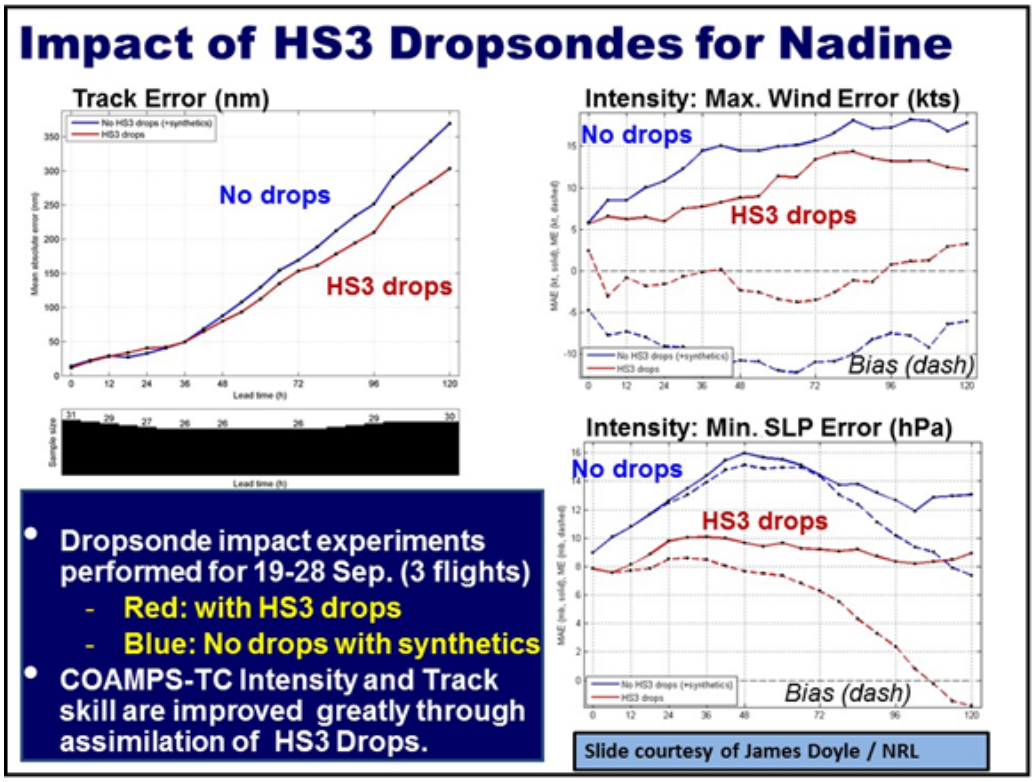

Figure 2.18. Slide provided by James Doyle demonstrating the significant positive impact on forecasts of Hurricane Nadine resulting from the assimilation of HS3 GH dropsonde data (drops) within the COAMPS-TC model. Results are shown for track error and intensity as reflected by both the maximum wind error and minimum sea level pressure (SLP) error.

A preliminary analysis of the impact of dropsonde observations collected during three flights into Hurricane Edouard in 2014 on COAMPS-TC forecasts was also presented at the 2014 Fall American Geophysical Union Meeting (Doyle et al. 2014). The track forecasts were again significantly improved but the impact on intensity forecasts was more mixed. The intensity forecasts were improved for two of the three flights but degraded on the other. In the case where the intensity forecast was not improved, the GH had a longer transit and spent more time over the center of the storm and less time sampling the environment. This suggests that

the majority of the improvement in forecast intensity within COAMPS-TC came from the observations of the broader environment around the storm. Interestingly, this is in contrast to the HWRF-HEDAS results where more benefit seemingly came from the observations nearer the storm center. This could reflect limitations associated with not incorporating dropsonde data into the HWRF-HEDAS boundary conditions but also highlights how observational data impacts can vary between different modeling systems.

Overall, it is significant that GH dropsonde observations have demonstrated potential to positively impact multiple independent modeling systems, though possibly in different ways. An analysis of the impact of the assimilation of different observation types, including satellite data, into the COAMPS-TC model showed that, on a per-observation basis, the dropsonde observations had the largest impact aside from synthetic observations. This is again a very positive result regarding the potential impact of GH observations.

2.1.5 External Remote Sensor Impact Studies

While the HIWRAP Doppler precipitation radar was flown in all of the SHOUT missions, processing of the data is more complex, and results were not available to the analysis teams in time for them to complete investigations into its impact. Some preliminary insight into its possible benefit can be gained from studies led by Jason Sippel using earlier data collected during the NASA Genesis and Rapid Intensification Processes (GRIP, Braun et al. 2013) experiment.

Analysis of HIWRAP data collected during Hurricane Karl in 2010 was presented by Sippel et al. (2014). This work followed an earlier analysis of simulated data from the same system by Sippel et al. (2013). The analysis of real data employed an ensemble Kalman filter (EnKF) within the Weather Research and Forecasting (WRF) Model (version 3.1.1) to compare the potential utility of assimilating HIWRAP Doppler velocity (Vr) and Doppler-derived velocity-azimuth display (VAD) wind profiles (VWPs). The Vr measurements are the most direct and highest resolution wind speed estimates and are consistent with approaches used with the existing tail Doppler radars, while the VAD method (Tian et al. 2015) estimates horizontal wind speed components through a sinusoidal fit to a full instrument scan as a way to minimize the impact of noisiness in Vr.

Data were obtained from sampling over a period of roughly 14 hours (with 20 eye overpasses) during the storm's motion across the Bay of Campeche on 16-17 September 2010. Use of the data was complicated by several issues associated with the fact that this was one of the first deployments of the instrument. Most notably, only the inner (30° incidence) beam was available. Use of the 40° incidence outer beam was assumed in Sippel et al. (2013) since those data have a smaller contribution from the vertical component of the wind speed which has less value for improving forecasts and analyses. Further issues included enhanced levels of noise in the Doppler measurements that required heightened levels of quality control and more data rejection.

Overall, the results showed that assimilation of either Vr or VWPs had a significant positive impact on the EnKF analysis, and forecasts initialized from the analysis were much more accurate than forecasts without assimilation (Sippel et al. 2014). The analyses were able to accurately estimate Karl's observed location, maximum intensity, size, precipitation distribution, and vertical structure. An example of the data's impact on forecasts from one model cycle is shown in Figure 2.19. Clear improvements in all quantities are observed relative to model runs without any data assimilation (NODA) or the control (CTRL) experiment without assimilation of HIWRAP. Generally, the forecasts initialized from the VWP-assimilating analyses performed slightly better than those from the corresponding Vr-assimilating analyses. The latter results were less accurate than for the simulated results in Sippel et al. (2013) likely because of lack of the outer beam data and increased real instrument noise.

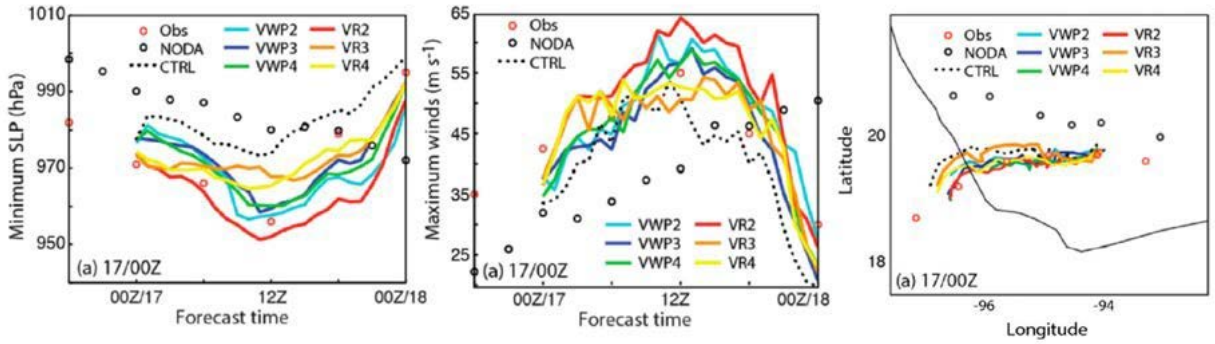

Figure 2.19. Results extracted from Sippel et al. (2014) highlighting the impact of assimilation of HIWRAP data from the GH on Hurricane Karl in 2010. The panels, taken from Figures 16-18 in the paper, show the corresponding impacts on minimum pressure (left), maximum wind (center), and track (right) for one forecast run initialized at 00Z on the 17th. The red circles denote the observed (Obs) results, the black circles correspond to no data (NODA) assimilation, the dotted black trace to the control (CTRL) run that assimilated just position and intensity information, and the color traces to additional assimilation of HIWRAP data in different Velocity-Azimuth Display Wind Profiles (VWP) experiments described in Sippel et al. (2014).

More experiments are again needed with the HIWRAP data collected during SHOUT, but the results are again very encouraging with respect to the potential for additional forecast improvement from remotely-sensed observations from the GH. Processed HIWRAP data are now becoming available and more analysis is clearly justified.

2.2 Global Tropical Cyclone OSE Results

The potential impact of GH observations on forecasts from global models are of equal interest to SHOUT as the results using the regional HWRF model. Global models, like the GFS, are a key resource used by forecasters in assessing potential storm impacts, particularly at longer lead times. These models rely heavily on the assimilation of satellite-based observations and the accuracy of their forecasts could be degraded were there any gap in the availability of these data. For these reasons, global tropical cyclone OSEs were an important element of the SHOUT analyses.

The global modeling studies evaluating the forecast impact of GH data all utilized versions of the operational GFS model. Investigations into GFS-based forecasts of hurricanes and tropical storms were conducted by both ESRL/GSD and NCEP/EMC. As noted previously, results from the GFS model also provide the boundary conditions for many of the HWRF model runs described previously.

2.2.1 GFS Hurricane Studies at ESRL/GSD

An initial investigation into the impact of GH observations on GFS forecasts of hurricanes was performed by Andrew Kren at ESRL/GSD under the leadership of Lidia Cucurull with a primary focus on evaluating the impact of the observations in the event of a gap in normal satellite coverage. This study used the version of the model that was operational as of May 2016 (January 2015 version) with horizontal resolution of T1534 (~13 km) for 0-10 day forecasts, and T574 (~34 km) for 10-16 day forecasts. This version used an ensemble Kalman filter-variational hybrid data assimilation (Wang et al. 2013) system based on three-dimensional variational data assimilation (Wu et al. 2002; Kleist et al. 2009). The most significant change to the operational model since that time has been to the data assimilation procedures. The current operational model now uses a 4-D Ensemble Kalman Filter approach. A complete description of this work is contained in a manuscript (Kren et al. 2017) submitted to Monthly Weather Review.

The study specifically examined the impact of dropsonde observations from the GH on forecasts of Hurricane Matthew following the first SHOUT flight on 5-6 October 2016. The experiments evaluated the impact of both full vertical resolution and the normal, reduced-resolution, dropsonde data. In the GFS studies, as with the operational HWRF analyses, the reduced resolution dropsonde data differed from the traditional TEMP DROP message content. While the TEMP DROP messages report data at specified mandatory and significant levels in the atmosphere, the observations here were binned to 38 equally spaced vertical levels from 100-1000 hPa, approximating the vertical resolution of the TEMP DROP messages. More importantly, the observations here incorporated position information at each vertical level as opposed to the TEMP DROP messages which only contain the launch position.

Four distinct experiments were conducted using the normal, reduced-resolution dropsonde data to evaluate their impact when added both to the complete existing observing system, and when added to a reduced observing system resulting from a gap in polar satellite coverage. The control (CTL) experiment included assimilation of all conventional observations in the absence of the GH. To simulate a gap in satellite coverage resulting from the loss of the Suomi NPP satellite, an additional experiment (noNPP) was conducted where ATMS and CrIS sounding data from the NPP satellite were withheld. GH dropsonde data were then added to both of these scenarios.

The dropsonde data were assimilated into the GFS over three cycle times from 06Z through 18Z on 5 October. The metrics evaluated included the traditional measures of forecast track

and intensity error as well as broader regional assessments of the impact on multiple atmospheric model fields. The regional assessment focused on a box extending between 25-35° N latitude and 73-83° W longitude. This region was defined based roughly on the spread in the model-based, forecasted position of Matthew on 8 October, two days after the first flight.

The impact of the GH dropsonde observations on the forecasted track and intensity errors at a lead time of 96 hours is summarized in Figure 2.20. The results demonstrate a reduction in the track error from addition of the dropsondes in both cases, but little change in the forecasted intensity as reflected by the central pressure and maximum wind speed. The reduction in the track error is approximately 10% but was not found to be statistically significant given the number of cycles analyzed. Interestingly, withholding the satellite data had little negative impact in this case. The lack of impact of the GH observations on the intensity forecast is not surprising for this flight since the sampling was entirely of the upstream environment while the center of the storm was near the Bahamas and inaccessible to the GH.

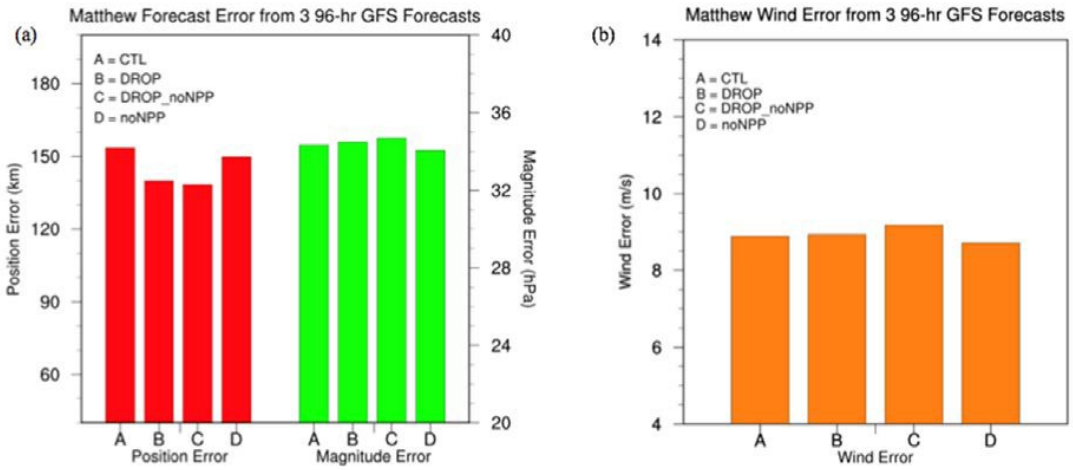

Figure 2.20. Impact of GH dropsonde observations (DROP) on forecasted track and intensity errors for Hurricane Matthew (2016) within the 2015 operational GFS model. The impact is shown relative to assimilation of both all conventional observations (i.e., the control (CTL) model) and for the case of a gap in Suomi NPP satellite data (noNPP). Results represent the forecast errors at 96 hours averaged over the three model cycles in which GH observations were available. Graphic provided by the Global Observing Systems Analysis (GOSA) Group as appearing in Kren et al. 2017.

To more broadly evaluate the impact of the GH dropsonde observations, their impact on anomaly correlations and root-mean-squared error (RMSE) for various model forecast fields at multiple vertical levels was computed both for the defined southeast region and over the continental United States (CONUS). Results summarizing the impact on the 500 hPa height and sea level pressure over the SE domain are shown in Figure 2.21 for the case of the full conventional observing system. Addition of the dropsonde data results in notable improvements in the anomaly correlation and reduction of the forecast error in both fields at longer lead times. Improvements are again on the order of 10% and were found to be statistically significant for these parameters. The corresponding results for addition of the

dropsonde observations in the event of a gap in the satellite observing system are shown in Figure 2.22. The dropsondes are observed to have a comparable positive impact, and the skill improvements and error reductions were again statistically significant.

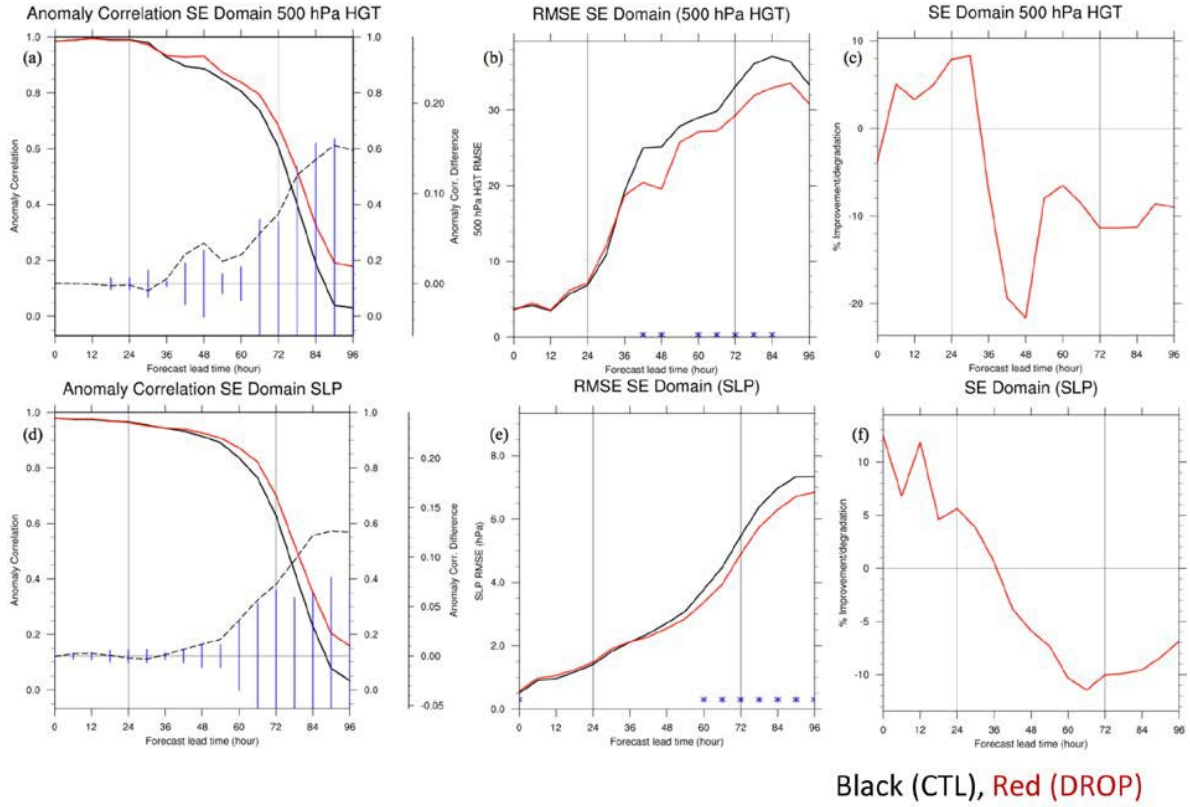

Figure 2.21. Impact of assimilation of Global Hawk dropsondes (DROP) versus the control (CTL) model on anomaly correlation scores and root mean squared error (RMSE) from GFS forecasts during Hurricane Matthew relative to the full observing system. The left panels show the anomaly correlation, the center panels display the RMSE, and the right panel shows the percent change in the RMSE as a function of forecast lead time. Upper panels are for the 500 hPa height field (HGT) and lower panels are for the sea level pressure (SLP). The results reflect averages from three model cycles and over the southeast (SE) verification region. In the anomaly correlation plots, differences (dashed lines) outside of blue vertical lines are significant at the 95% confidence level. Blue stars in the RMSE plots indicate lead times that are statistically significant at the 95% level. Graphics provided by the GOSA Group as appearing in Kren et al. 2017.

The relative impact of the dropsonde observations on the RMSE in several additional model fields over the SE verification region at lead times up through 96 hours is further displayed in Figure 2.23 for both the satellite gap scenario and for addition to the full current observing system. The reference for these results and the figures above was the European Centre for Medium-Range Weather Forecasts (ECMWF) analysis, which did incorporate GH observations. Reduced errors are observed in nearly all the fields considered, emphasizing that the positive forecast impacts are largely independent of the chosen validation variable. Improvements appear greatest in the height and wind fields with the largest reductions approaching 12%. On average, the impact is slightly more positive in the satellite gap scenario. The increased impact

of the dropsondes in the gap scenario relative to the full observing system is most prominent for the upper level observations. This highlights the potential unique capability of the GH to provide otherwise unavailable upper-atmospheric measurements in the absence of key satellite observations.

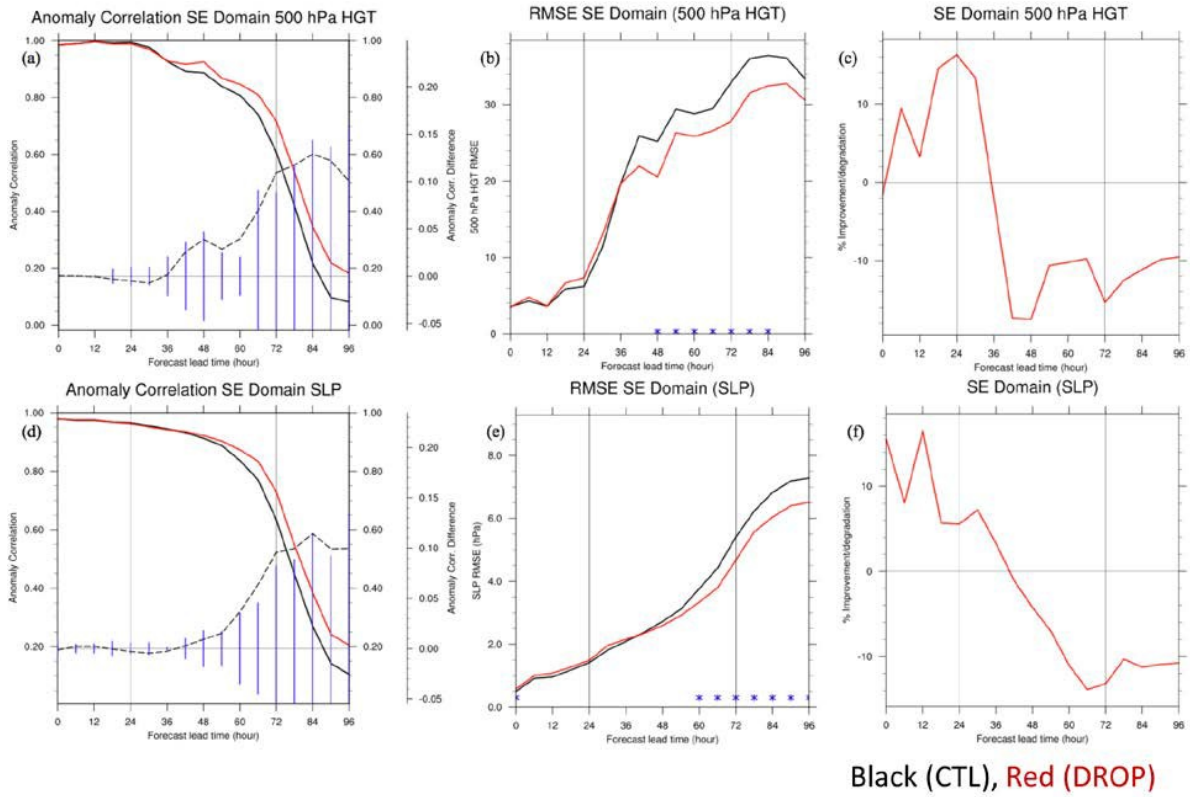

Figure 2.22. Same as Figure 2.21 but for the satellite gap scenario. Graphics provided by the GOSA Group as appearing in Kren et al. 2017.

Of particular interest is the impact of the dropsonde observations on forecasted precipitation in the southeastern United States. The change in the forecasted 4-day accumulated precipitation and associated errors are shown in Figure 2.24 relative to the control case assimilating all conventional observations. The improvement in forecast track results in a notable improvement in the forecasted precipitation, particularly over and offshore of South Carolina where the heaviest precipitation is shifted further onshore, in agreement with the observations. Similar improvements were also observed (not shown) for the satellite gap scenario.

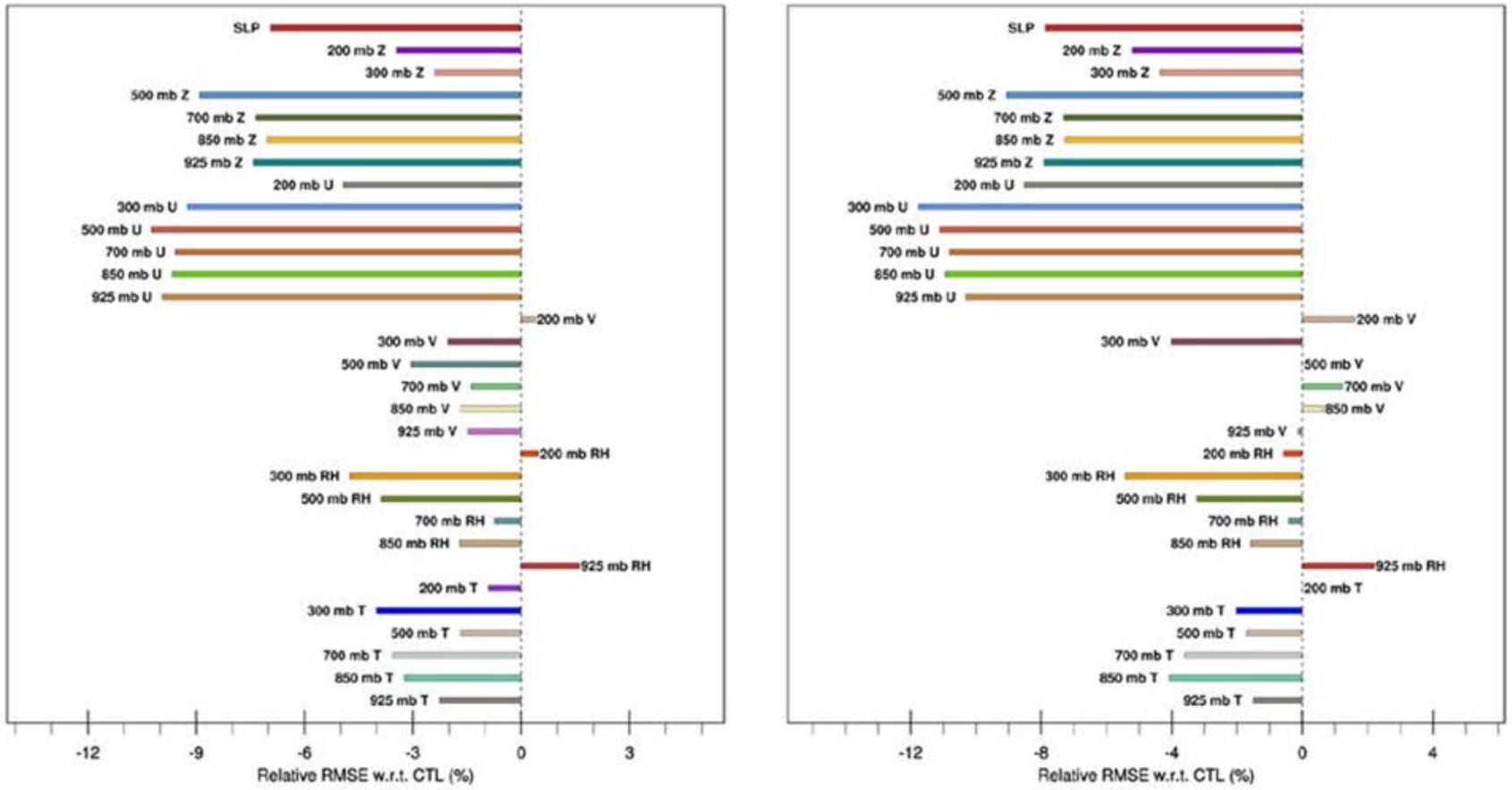

Figure 2.23. Graphical illustration of the relative change in forecast root mean squared error (RMSE) in multiple forecast fields (e.g., sea level pressure (SLP), height (Z), wind speeds (zonal, U, and meridional, V, components), relative humidity (RH), and temperature (T)) resulting from inclusion of the GH dropsondes. The left panel is for the full observing system and the right panel is for a gap in satellite coverage. The results represent averages over the southeast verification region and all lead times up to 96 hours for the three forecast cycles. Negative values represent reduction in forecast error. Graphics provided by the GOSA Group.

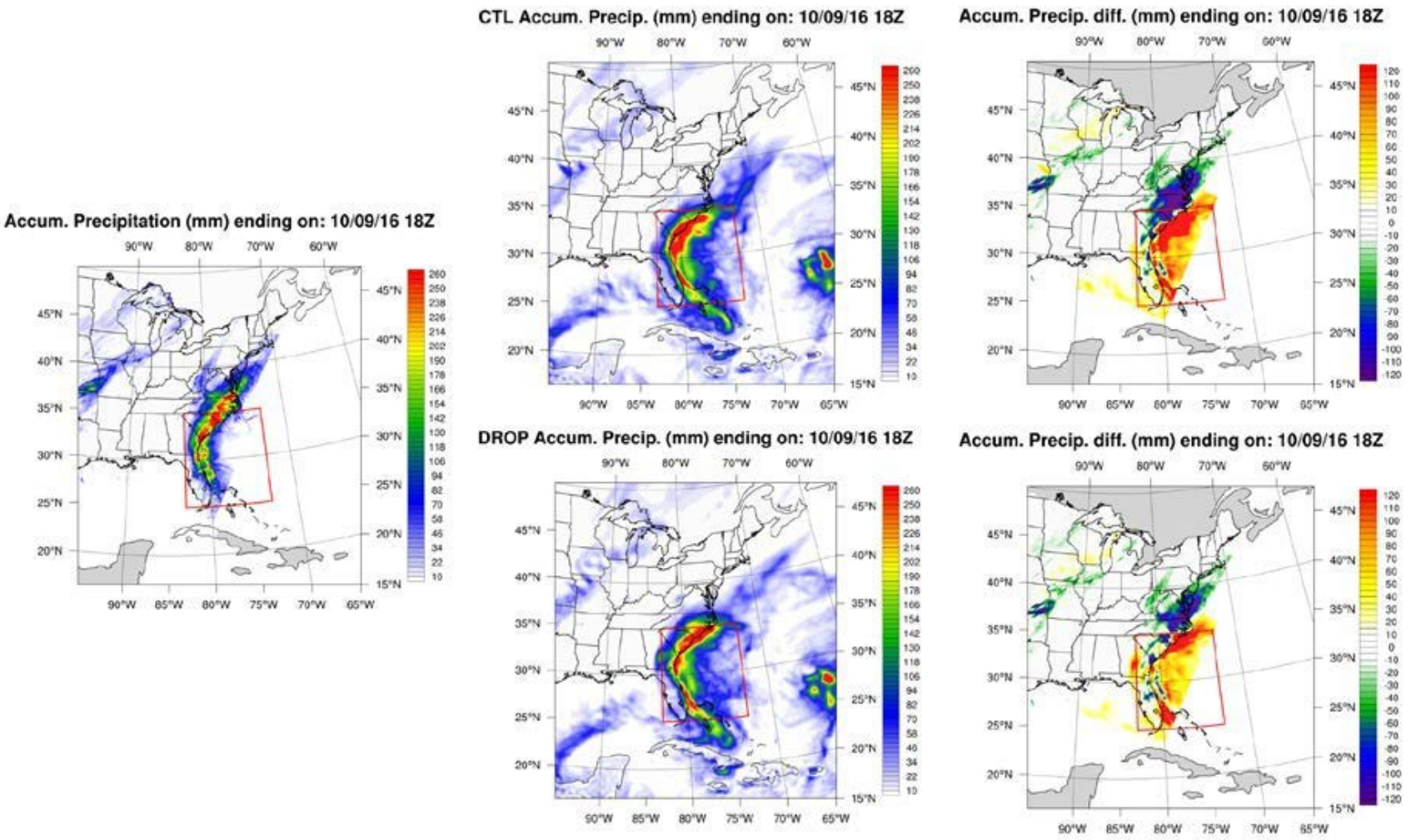

Figure 2.24. Impact of GH dropsondes on precipitation forecasts from GFS during Hurricane Matthew. Results are shown for 4-day accumulated precipitation ending at 18Z on 9 October. The observed precipitation is shown at the left; the forecast precipitation is shown in the center for the control run assimilating all standard observations (CTL, top) and with addition of the GH dropsonde data (DROP, bottom); and the corresponding errors in the forecast precipitation are shown at the right. Graphics from the GOSA Group.

To further quantify the impact of the dropsondes on forecasted precipitation, equitable threat scores for various thresholds of 24 h accumulated precipitation were computed for different forecast periods and the results, averaged over the CONUS, are presented in Figure 2.25. Inclusion of the dropsondes resulted in a more skillful forecast for all precipitation thresholds and forecast hours both with and without a gap in satellite coverage. The impacts are greatest at the longest forecast leads. Given the large societal impact of precipitation resulting from Matthew, these results, consistent with what was observed with HWRF, are especially positive with respect to the importance of the dropsonde measurements.

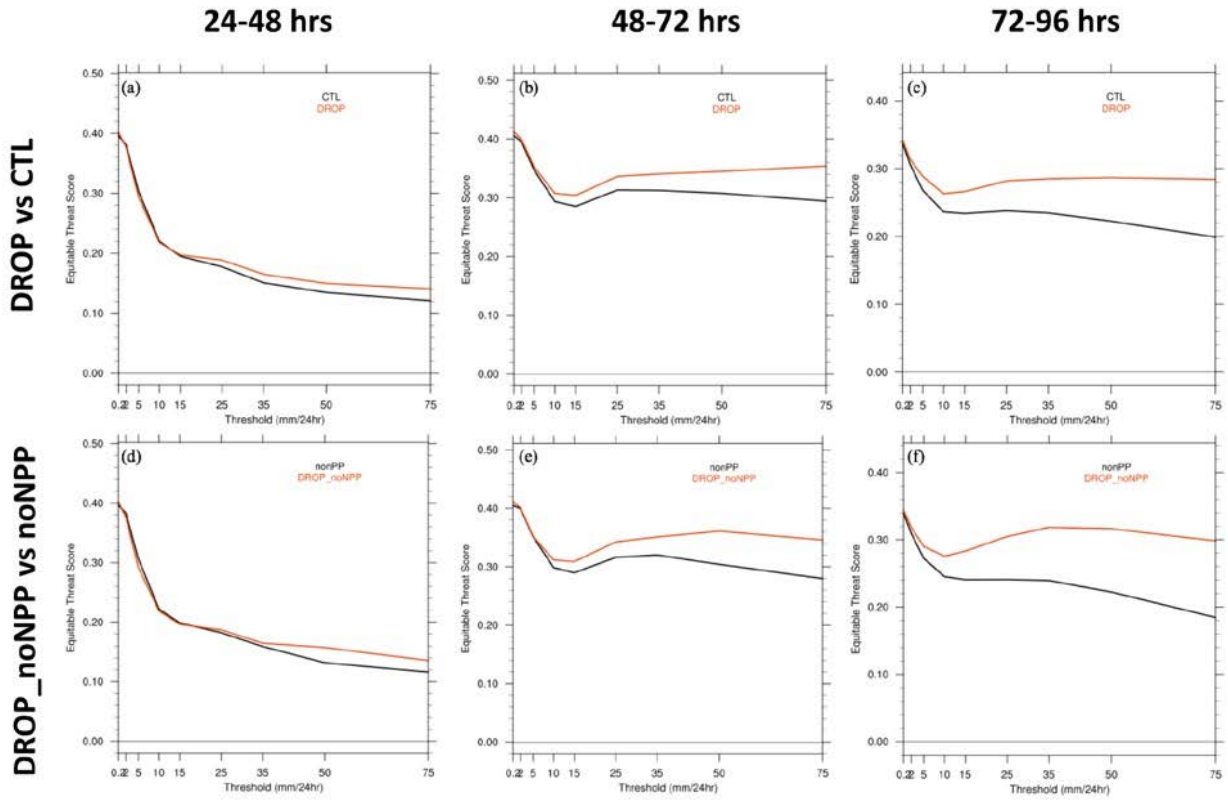

Figure 2.25. Equitable Threat Score averaged over all assimilation cycles and over the CONUS as a function of various thresholds of 24-hour accumulated precipitation (0.2, 2, 5, 10, 15, 25, 35, 50, and 75 mm $24h^{-1}$) from forecast hours (left) 24 to 48, (center) 48 to 72, and (right) 72 to 96 hours. The upper panels are for the comparison with assimilation of all conventional observations (i.e., control [CTL]) while the lower panels are for the gap scenario (noNPP). The red traces reflect addition of the GH dropsondes (DROP) compared with the black traces without. The horizontal black lines on the plots represent no skill in the forecast. Graphics provided by the GOSA Group as appearing in Kren et al. 2017.

Errors in the multiple forecast fields were also examined when averaged over the CONUS and globally to determine the larger scale impact of the observations. The statistics (not shown) generally reflected neutral results or small degradations. While the positive impacts were largely constrained to the immediate region of the storm, the observations did not significantly harm the forecasts globally, particularly in the case of a gap in satellite coverage.

An additional experiment evaluated the impact of assimilating the full vertical resolution dropsonde data from the GH in addition to the full conventional observing system. The results for the track and intensity forecasts at 96-hr lead are shown in Figure 2.26. Addition of the full resolution data results in a major reduction in the track error. The reduction is approximately 30% over the control forecast assimilating all traditional observations and about 25% over the previous results obtained assimilating the normal, reduced resolution, GH dropsonde data. There is again little impact on the intensity forecast as reflected in the central storm pressure, but there is a slight reduction in the error in the forecasted wind speeds. These results, in this case, suggest tremendous potential future value in incorporating the full resolution data available from the dropsondes. This same level of success, however, was not obtained in all experiments using the full resolution data, and there may still be some issues with the ability of the GFS model to uniformly handle such high-vertical-resolution observations.

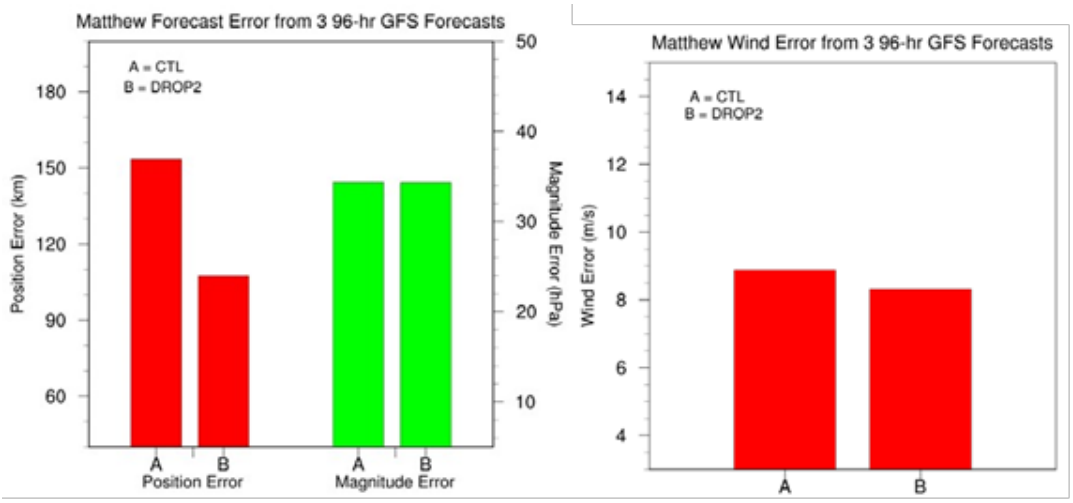

Figure 2.26. Impact of full-resolution GH dropsonde observations (DROP2) versus the control (CTL) model on GFS-forecasted track and intensity errors for Hurricane Matthew. Results are again shown at 96-hours relative to assimilation of all standard satellite observations. Graphics provided by the GOSA Group.

While for one storm only, the findings are quite positive with respect to the ability of GH dropsonde observations to improve forecasts of tropical cyclones within operational global models, particularly in the presence of a gap in key satellite observations. The benefits were reflected in forecasted track and in a variety of metrics for many different forecast fields. The improvements are largely constrained to the regions directly affected by the storm, but any detrimental effects in other regions were small and limited in extent. A very preliminary investigation into the impact of HAMSR retrievals on GFS forecasts of Matthew yielded mixed results with insignificant changes to the forecast skill. Further work exploring the treatment of HAMSR data is still required.

2.2.2 GFS Hurricane Studies at NCEP/EMC

Investigation of the impact of GH dropsonde observations on tropical cyclone forecasts within

the operational GFS model in the presence of the full current observing system was expanded significantly at NCEP/EMC. The results, described below, are extremely positive and reflect some of the strongest arguments for potential future utilization of GH-type observations in an operational program.

The GFS-based studies conducted at NCEP/EMC utilized the 2017 version of the model that went into operations in the middle of the year. The experiments incorporated the GH dropsonde observations collected in all flights during the 2014 HS3 and 2016 SHOUT hurricane campaigns which encompassed seven named storms. The impact of the observations were evaluated relative to assimilation of all conventional and reconnaissance observations utilized in normal operations. The dropsondes were not included in the original operational runs at the time of the flights. The work was conducted by Jason Sippel and Kate Howard under the leadership of Vijay Tallapragada.

As for the operational HWRF analysis described in Section 2.1.2, the GH dropsonde data was obtained directly from the TEMP DROP messages submitted in real time through the GTS. In the TEMP DROP messages, measurements were reported only at specified mandatory and significant levels and not at the full measurement rate. The data also contained only the launch location and not any information on absolute position during the descent.

The overall impact of the observations from 2016 on the GFS track forecasts of all concurrent Atlantic basin storms is shown in Figure 2.27. The results reflect a substantial positive impact on the track forecast at lead times beyond about 36 hours. The relative skill improvement peaks at around 14% for 72-hour forecast lead but exceeds 10% for leads greater than 48 hours. These overall improvements were found to be just on the edge of statistical significance at the 95% level as reflected by the error bars in the figure.

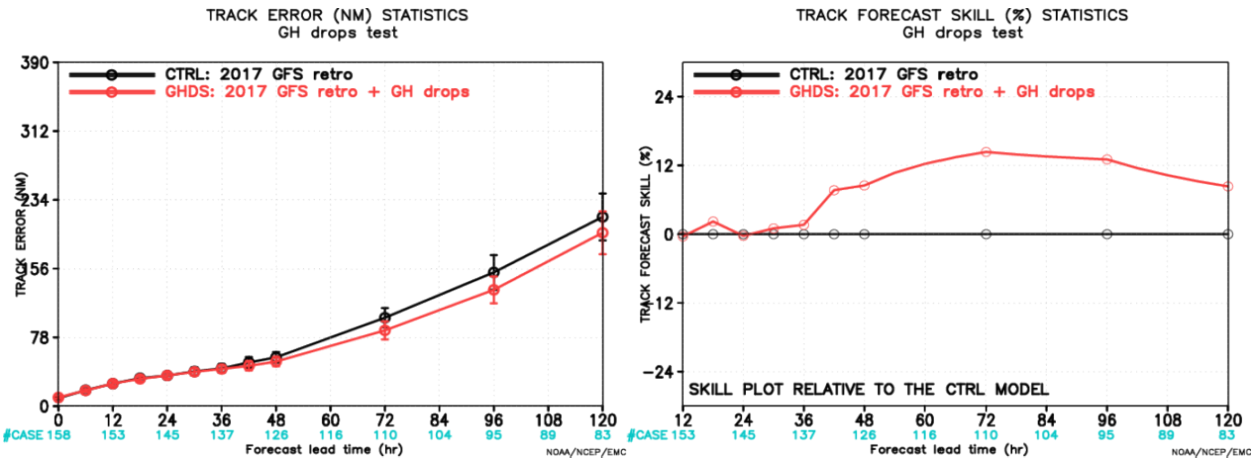

Figure 2.27. Average impact of Global Hawk dropsonde measurements (GH drops) on track errors within the 2017 operational GFS model at NCEP/EMC. The results represent an average over the forecast cycles spanning all storms sampled by SHOUT in 2016. The left panel shows the track errors with (i.e., GHDS, red) and without (i.e., the control (CTRL) model, black) the GH dropsondes while the right panel shows the corresponding change in skill. Graphics provided by Jason Sippel.

Examination of the GH dropsonde impact on individual 2016 storm forecasts shows dramatically greater benefit to the track forecasts at extended lead times. Results obtained for model cycles during Hermine and Gaston are shown in Figure 2.28. For each storm, highly significant track improvements in excess of 20% are observed for extended periods in the forecasts. Forecasts during Hermine exhibited high track uncertainty even with extensive reconnaissance from the NOAA WP-3D and Air Force WC-130J aircraft. Addition of the GH dropsondes led to strong improvements throughout the forecast period. Forecast gains during Gaston reflect improvements in the absence of any other reconnaissance data.

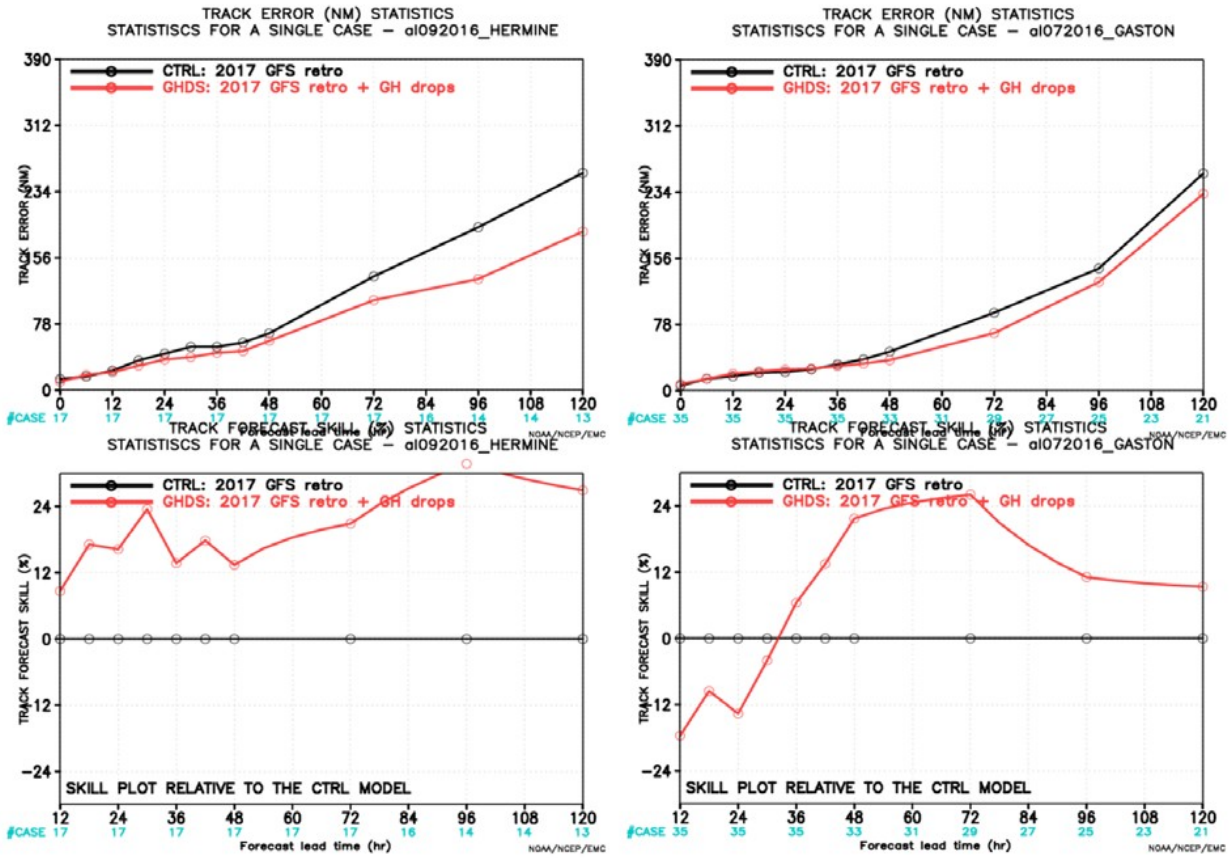

Figure 2.28. Impact of GH dropsonde (GH drops) measurements on forecasted track errors within the 2017 operational GFS model for Hermine (left) and Gaston (right). The results are averages over the forecast cycles concurrent with the flights during the individual systems. The upper panels show the track errors with (i.e., GHDS, red) and without (i.e., control (CTRL) model, black) the GH dropsondes while the lower panels show the corresponding change in skill. Graphics by Jason Sippel.

Remarkably, the GH observations of the Atlantic storms in 2016 also led to notable improvements in storm forecasts in the Pacific Ocean. The average impact of the observations on track forecasts of concurrent storms in the eastern and western Pacific basins are shown in Figure 2.29. Large track forecast skill improvements are observed for 72- to 120-hr forecasts. The eastern Pacific improvement at 72 hours is statistically significant and the western Pacific

improvement at 96 hours is nearly significant as well. Investigation of the results suggested that the positive remote influences were not random and it was hypothesized that assimilation of the dropsonde data could be positively impacting the satellite bias correction within the model leading to broader scale benefits.

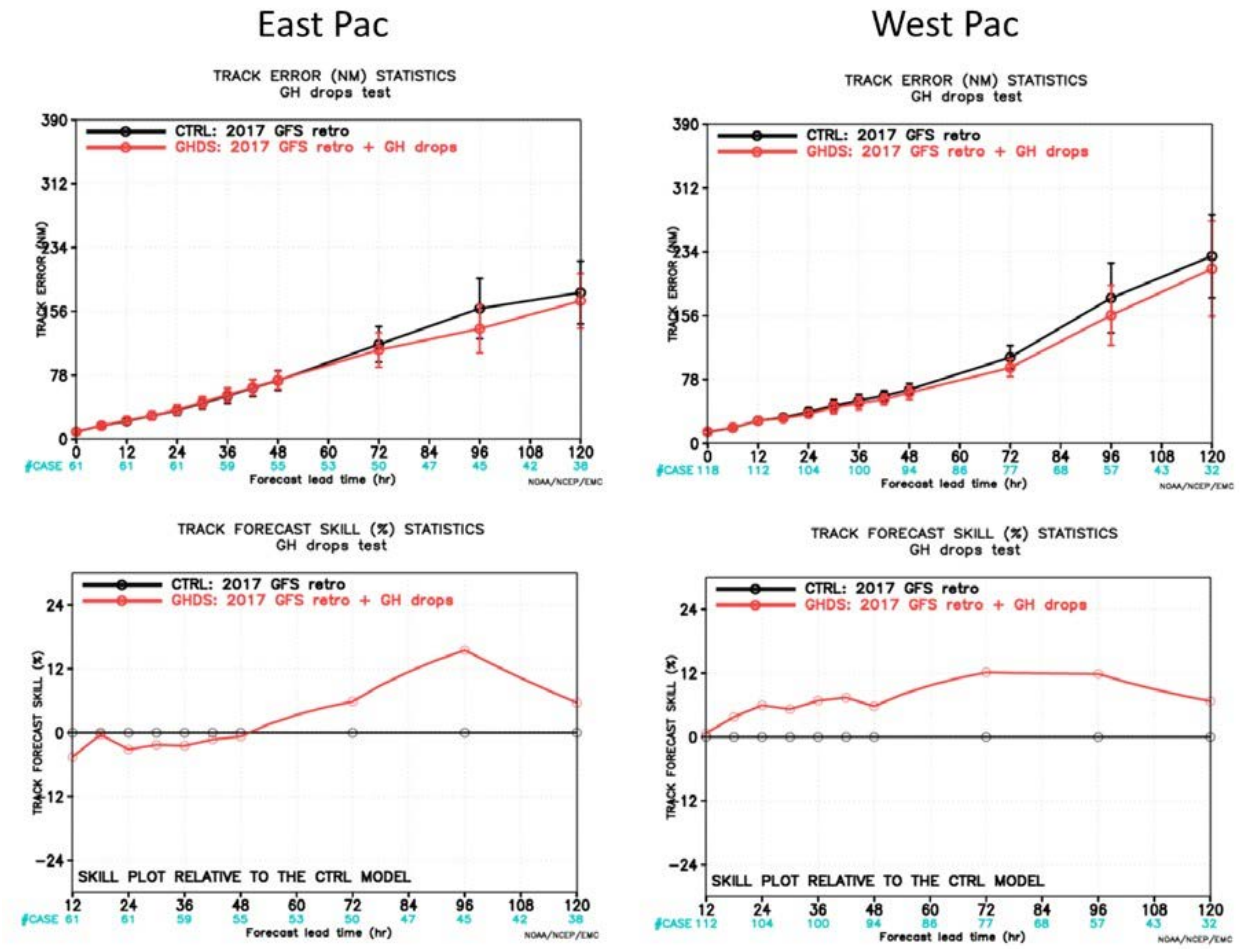

Figure 2.29. Impact of 2016 GH dropsonde (GH drops) observations in the Atlantic on concurrent storms in the Pacific. The upper panels show the track errors with (i.e., GHDS, red) and without (i.e., control (CTRL) model, black) the GH dropsondes while the lower panels show the corresponding change in skill. Graphics provided by Jason Sippel.

Use of the GH dropsondes from 2014 was complicated by an upper-level dry bias issue that affected all real-time dropsonde data (from manned and unmanned platforms) from 2010-2015. The problem, documented by Vömel et al. (2016), affected humidity measurements at temperatures colder than about -10°C. Experiments compared assimilating the dropsonde data as it was collected, with the bias, and by excluding all humidity measurements above 500 hPa. Slightly better results for the 2014 Atlantic storms were obtained assimilating the data with the bias, but the results were not significantly different. The biased moisture measurements in the Atlantic did, however, have a strong negative impact in the eastern

Pacific, further supporting the above assertion that the remote observational impacts were not random. Bias corrected data do now exist, but not in the operational TEMP DROP format utilized by the assimilation system.

Results combining the biased 2014 observations with the 2016 data reflecting the overall GH dropsonde impact on track forecasts in the Atlantic are shown in Figure 2.30. The improvement in skill is remarkably similar to that for 2016 alone, showing forecast gains at lead times from 42 hours forward. The peak improvement is about 15% at 72 hours and the improvement is statistically significant at the 95% level at 72 and 96 hours. The corresponding overall GH dropsonde impact on the forecast intensity error in GFS as reflected by the maximum wind speed was largely neutral throughout the forecast period (Figure 2.31). Jason Sippel, however, did not attribute much relevance to operational storm intensity forecasts obtained from the GFS model.

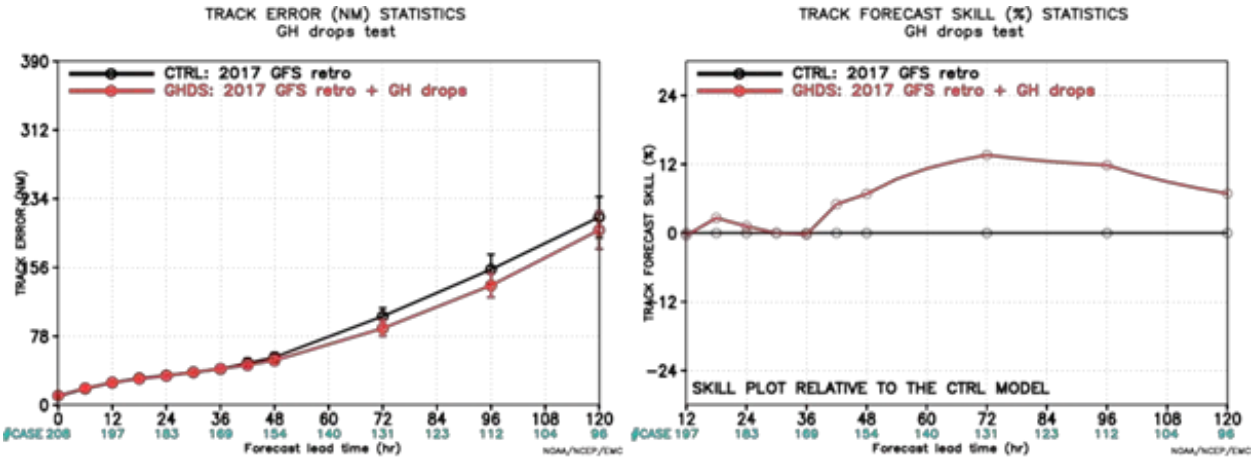

Figure 2.30. As in Figure 2.27, but for results averaged over the combination of flights from both 2014 and 2016. Graphics provided by Jason Sippel.

The results reported in this section are important for multiple reasons. First, they reflect the forecast benefit that could be obtained immediately from operational utilization of dropsonde measurements from the GH platform. Second, based on these results, the decision was made at EMC to remove the flag previously excluding the operational assimilation of GH dropsondes within GFS. This meant that GH observations collected in partnership with a NASA experiment in August 2017 were operationally assimilated in GFS. Finally, they represent significant excitement on the part of the modelers at EMC. In presenting these results in SHOUT briefings, both Jason Sippel and Vijay Tallapragada commented on the significant nature of the results relative to other forecast gains observed recently within the operational modeling system. With respect to the SHOUT objective of evaluating the potential impact of GH observations in the event of a satellite gap, while the operational results did not explicitly consider a possible satellite gap, the ability to obtain such notable forecast improvements on top of the full current observing system strongly suggests that the observations should have an important positive impact in the absence of elements of the satellite observing system.

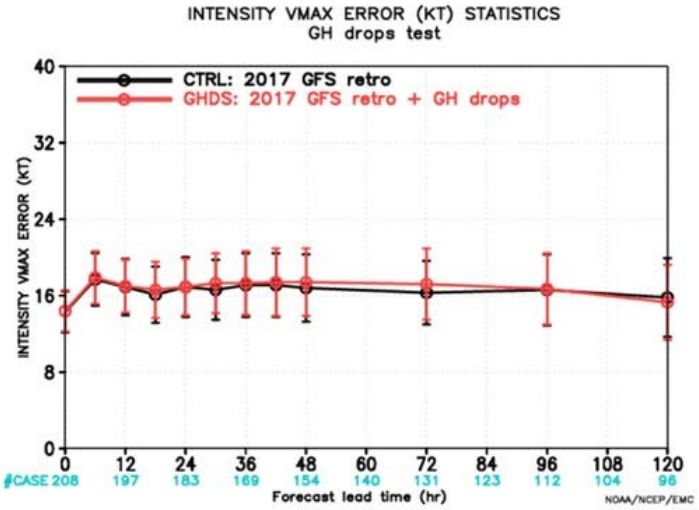

Figure 2.31. Corresponding results summarizing the impact of GH dropsonde observations from 2014 and 2016 on GFS forecasts of storm intensity. Graphic provided by Jason Sippel.

2.3 Pacific Storm OSE Results

The impact of GH observations on forecasts of high-impact weather events other than tropical cyclones is also important when assessing the potential merits of the platform. An additional investigation was conducted at ESRL/GSD by Andrew Kren under the leadership of Lidia Cucurull using the GFS model to determine the potential benefit of GH observations for high-impact Pacific wintertime storms like those sampled during the ENRR experiment. The modeling studies again used the same model configuration described in Section 2.2.1. The experiments focused on the impact of observations collected from the GH on 21-22 February 2016 in advance of a high wind and precipitation event in southern Alaska on 24 February. This event was selected because of its impact, the good performance of the GH dropsonde system during that flight, and the extensive coordination with other aircraft.

The experiments evaluated the impact of the GH dropsondes relative to forecasts obtained using the current complete operational observing system and when observations from the Suomi NPP satellite were withheld to simulate a gap in the satellite observing system. The GFS model was run for four cycles between 12Z on 21 February and 12Z on 22 February covering the period of available GH observations. The impact of the observations was again evaluated based on errors in a total energy measure and on multiple forecasted model fields evaluated over different spatial domains. For more direct comparison with the tropical cyclone studies, errors in the mean storm track and intensity were also evaluated. The "truth" or reference for the forecast assessments was taken to be the ECMWF analysis.

The primary region for the forecast assessment was centered on the southern coast of Alaska, corresponding to the greatest direct storm impact. The selected region was also the target region for the sensitivity calculations used in designing the GH flight track. The region is

outlined in Figure 2.32 along with a graphical representation of the magnitude of the forecast impact as a function of geographic location for the case where the GH dropsondes were added to the conventional observing system. The graphics reveal a notable reduction in the 2-day forecast errors within the region, particularly for the 200-mb height and sea level pressure fields. The greatest forecast improvements are generally centered within the target region with a mix of smaller improvements and degradations in other locations. The corresponding results obtained when the dropsondes were applied in the absence of Suomi NPP observations were generally similar in spatial extent but reduced in magnitude (not shown).

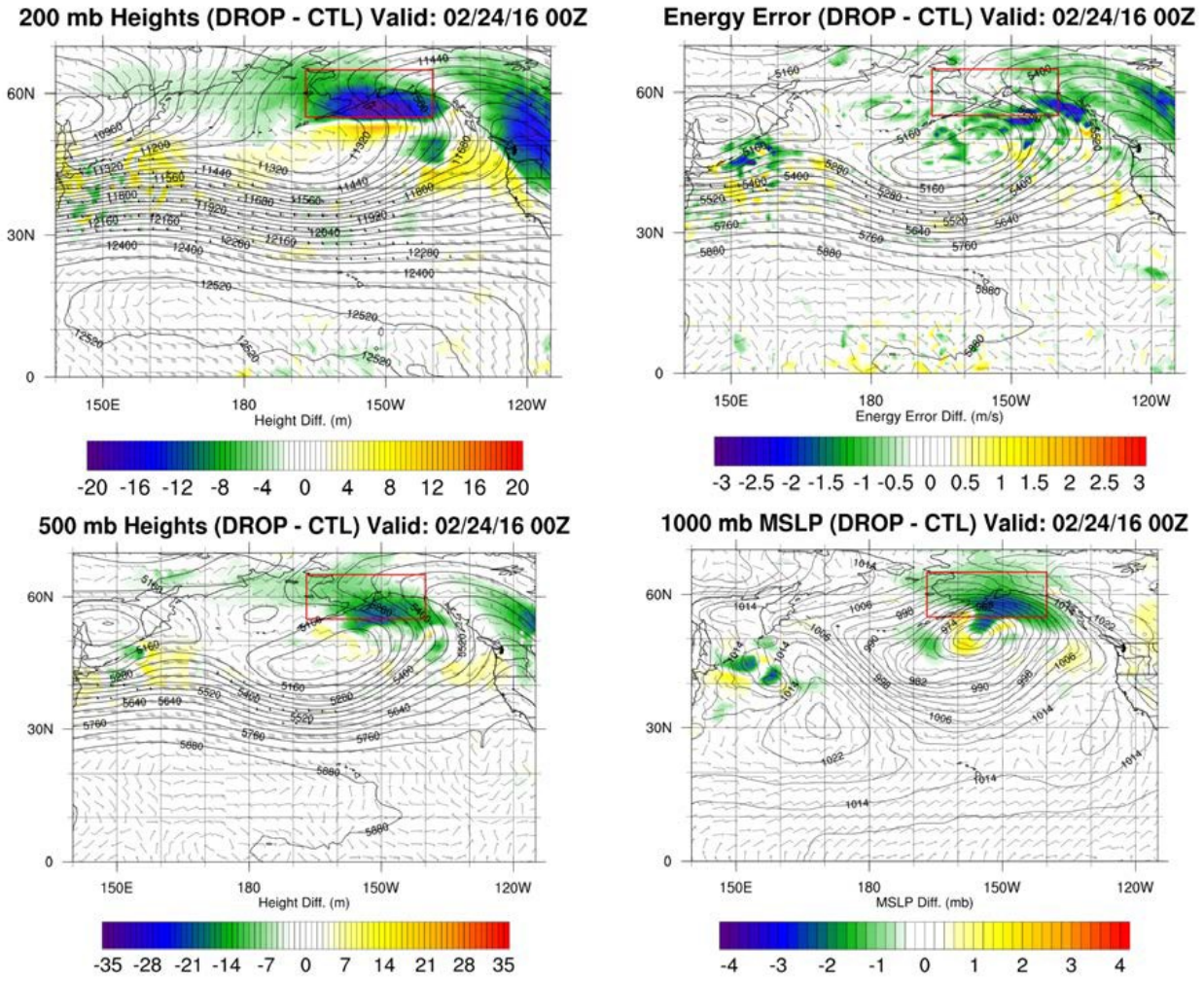

Figure 2.32. Impact of GH dropsondes (DROP) from the 21-22 February flight on GFS forecasts of the indicated fields valid at 00Z on 24 February. Results are shown as changes in the errors relative to forecasts assimilating all normal satellite observations (control, CTL). Negative (cold) values represent reductions in the errors by addition of the dropsonde observations. Graphics provided by the GOSA Group.

The observation impact on the anomaly correlation and RMSE for the 500-hPa height field and sea level pressure is shown in Figure 2.33 averaged over the target Alaska domain for the case of addition to the full conventional observing system. The results suggest general

improvement in the forecasts, particularly over the period from 24-72 hours. The nature of the improvement is broadly similar to that observed for Hurricane Matthew in Figure 2.21, but the benefit is smaller in magnitude. The RMSE improvement reach maximum values from 4-6% at lead times from 48 to 60 hours, corresponding to the occurrence of the highest impact weather. The improvements were not, however, found to be statistically significant at the 95% confidence level.

The corresponding results for the satellite gap scenario are shown in Figure 2.34. Improvements resulting from assimilation of the dropsonde data are observed over a larger portion of the forecast period, and the peak RMSE reductions ranging from 5-7% are greater than when the data were added to the full observing system. The improvements, while notable, were still not found to be statistically significant based on the limited sample size.

The broader impact of the GH observations was again investigated by comparing the RMSE in multiple model fields for forecasts with and without the GH data. The change in the RMSE values within the AK domain at forecast lead times through 96 hours is shown in Figure 2.35 for both the satellite gap scenario and for addition to the full current observing system. As for Hurricane Matthew, reduction in forecast errors results for nearly all variables and levels. The overall improvement is smaller than for Matthew, however, with reductions typically less than 4% for the full observing system and values just slightly greater on average in the event of a gap in satellite coverage. Not all individual fields, however, are improved by a greater amount in the event of a satellite gap. The greatest increase in observational impact in the gap scenario is often observed in upper level model fields where few other observations are available.

Forecast impacts were also evaluated over the CONUS and globally. General improvements observed over the Alaska domain did not extend uniformly to global scales, but the impact of the GH dropsondes was better in the event of a satellite gap. The impacts were largely neutral globally, but there was a small but significant degradation in the anomaly correlation for the 500-hPa height field at 84-96 hours when the dropsondes were added to the full observing system. The negative impact of the GH observations in this instance was found to be caused by poorer representation of the 500-hPa height field in the region of a storm outbreak over the southeastern United States. The degradation was not observed in the satellite gap scenario. Addition of the GH data in the absence of satellite observations produced almost universally neutral results for all fields outside of the AK domain. The fact that global results are not adversely affected is positive in this instance where the observations were specifically targeting benefits in one region.

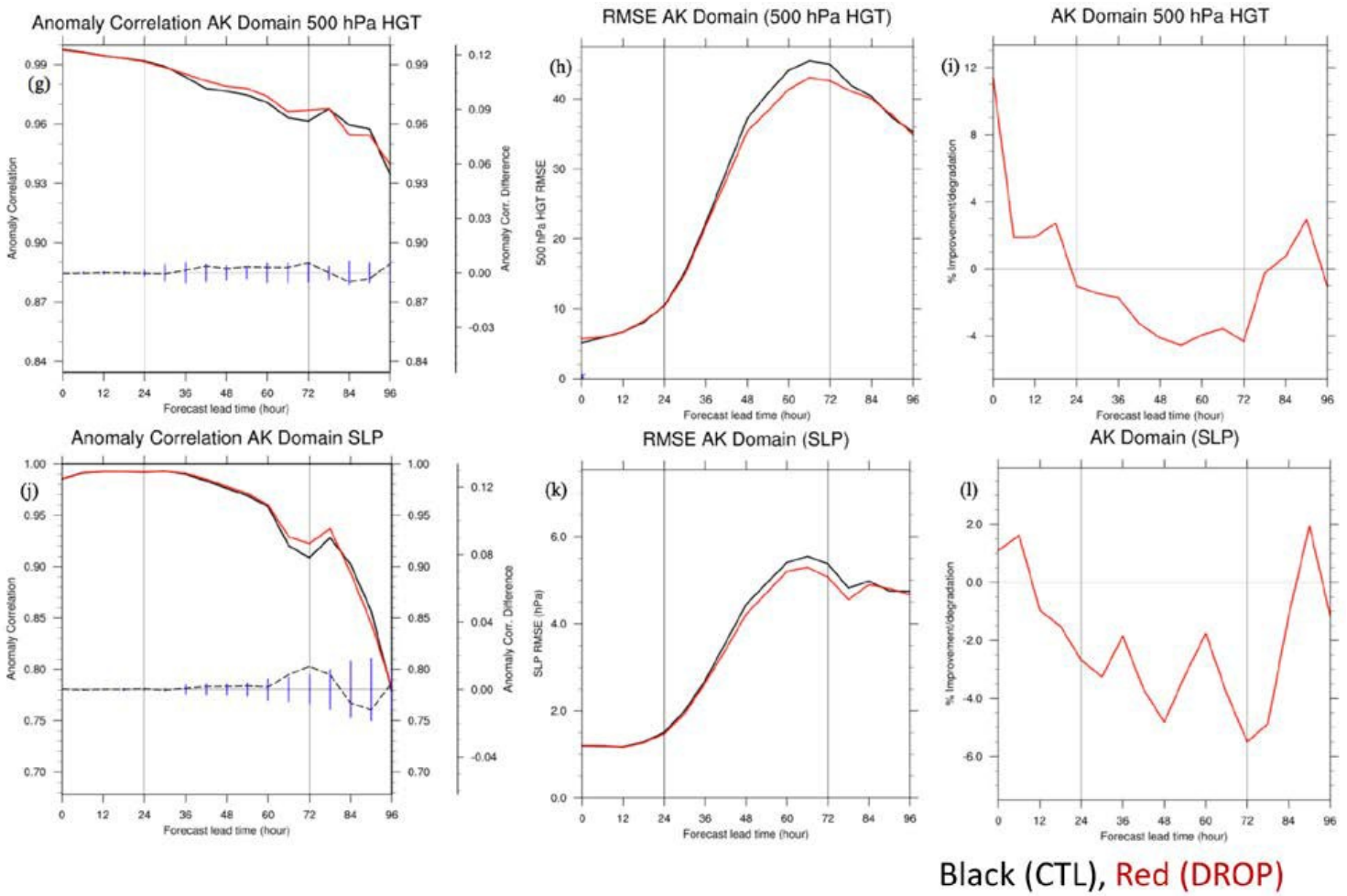

Figure 2.33. Impact of assimilation of GH dropsondes from 21-22 February on anomaly correlation scores and root mean squared errors (RMSEs) from GFS forecasts for the targeted Alaska region relative to the full observing system. The left panels show the anomaly correlation, the center panels display the RMSE, and the right panel shows the percent change in the RMSE as a function of forecast lead time. Upper panels are for the 500 hPa height field and lower panels are for the sea level pressure. The results reflect averages from four model cycles between 18Z on 21 February and 12Z on 22 February and over the verification region. The primary traces reflect inclusion of the dropsondes (DROP, red) and the control run without (CTL, black). In the anomaly correlation plots, differences (dashed lines) outside of blue vertical lines are significant at the 95% confidence level. Blue stars in the RMSE plots indicate lead times that are statistically significant at the 95% level. Graphics provided by the GOSA Group as appearing in Kren et al. 2017.

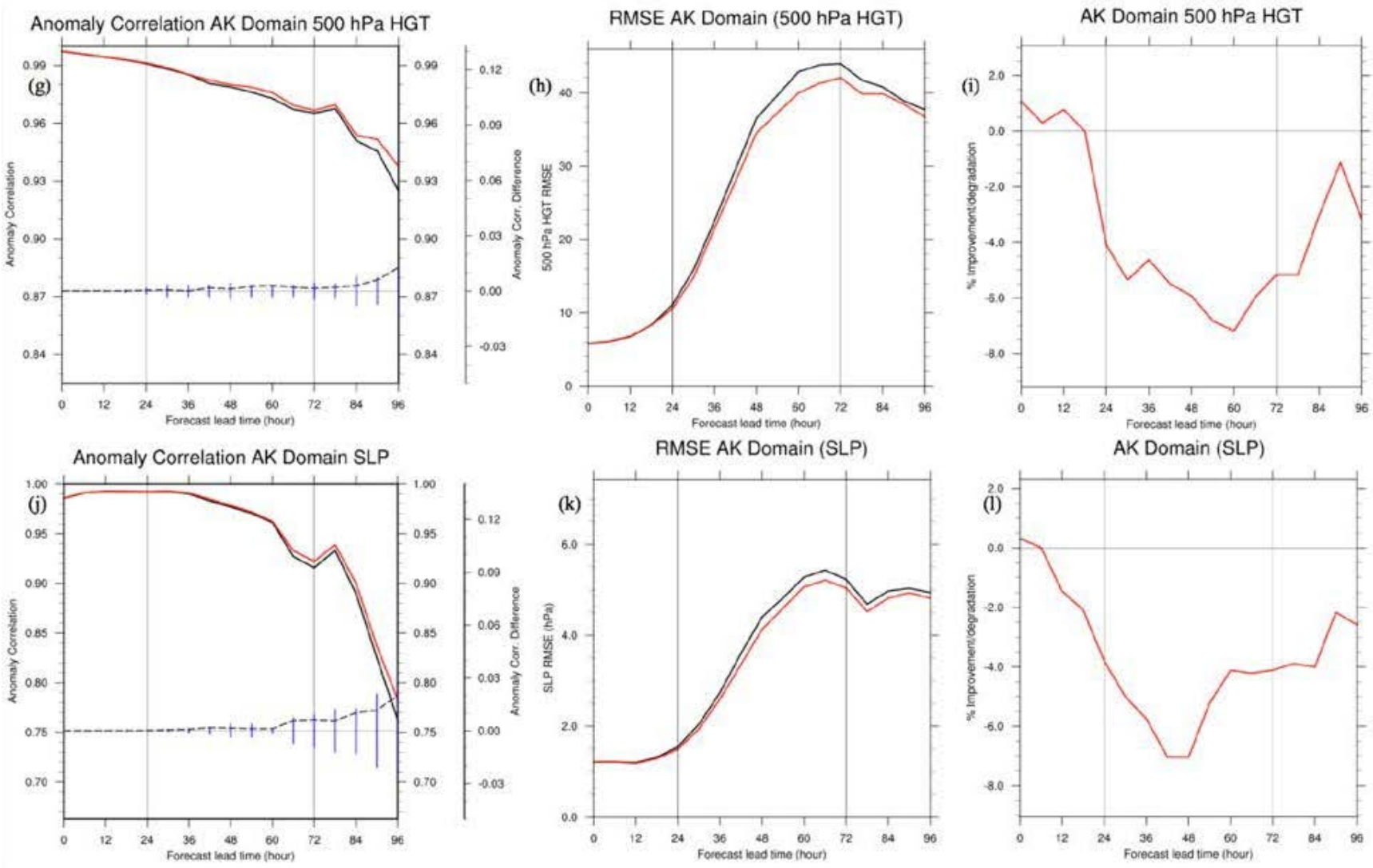

Figure 2.34. Global Hawk dropsonde (red) impact versus the control model (black) as in Figure 2.33, but in the event of a gap in satellite coverage. Graphics provided by the GOSA Group as appearing in Kren et al. 2017.

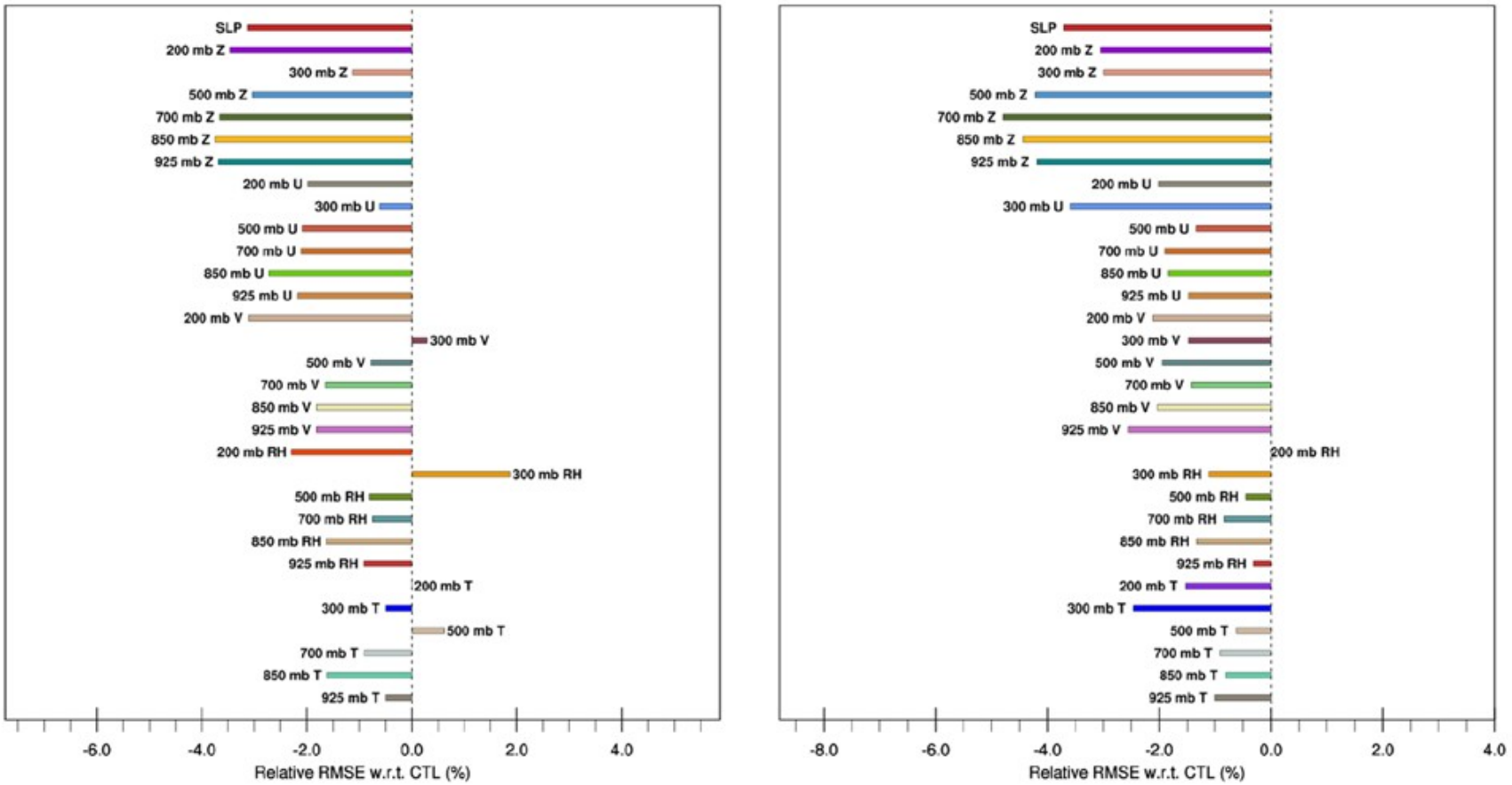

Figure 2.35. Graphical illustration of the relative change in forecast root mean squared error (RMSE) in multiple forecast fields (e.g., sea level pressure (SLP), height (Z), wind speeds (zonal, U, and meridional, V, components), relative humidity (RH), and temperature (T)) resulting from inclusion of the GH dropsondes with respect to the control (CTL) model. The left panel shows results for a full observing system and the right panel for the satellite gap scenario. The results represent averages over the Alaska verification region and all lead times up to 96 hours for the three forecast cycles. Negative values represent reduction in forecast error. Graphics provided by the GOSA Group.

In analogy with the tropical cyclone results, measures of the winter cyclone's forecasted track and intensity (as reflected in the central pressure) errors were computed and are displayed in Figure 2.36 for a lead time of 72 hours. The positive impact of GH observations in the presence of a satellite gap is again visible. Removal of the Suomi NPP data resulted in a notable increase in the position error which was then partially reduced (mitigated) through addition of the GH data. The intensity forecast of the storm actually improved slightly when the satellite data were withheld and addition of the dropsonde data had negligible impact.

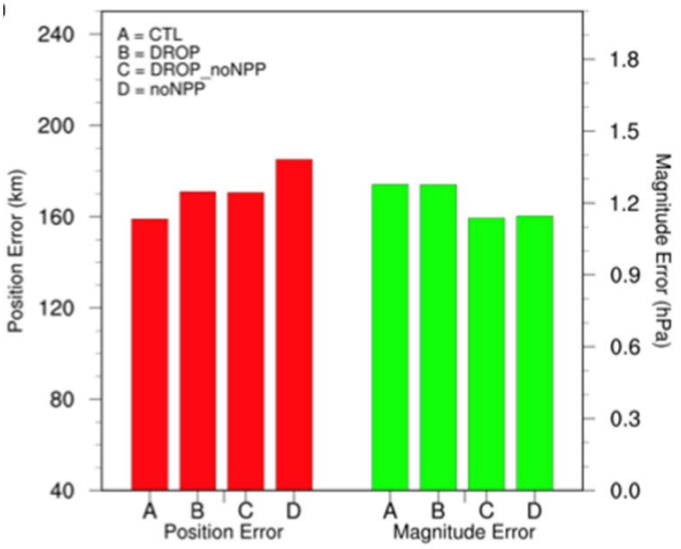

Figure 2.36. Impact of GH dropsonde observations from 21-22 February on GFS forecasts of the track and intensity of the Pacific cyclone that subsequently impacted southern Alaska on 24 February. Results are shown for addition of the dropsondes (DROP) to scenarios both with (noNPP) and without (control, CTL) a gap in satellite coverage as reflected by the legend. Graphic provided by the GOSA Group as appearing in Kren et al. 2017.

An additional study of the GH dropsonde observations from the ENRR campaign within the current operational GFS model is just now underway at NCEP/EMC. Very preliminary results examining the impact of the observations from a portion of the flights on 500-hPa anomaly correlations for 120-hr forecasts over the northern Pacific and Northern Hemisphere demonstrate a clear positive signal. These results are notable additions to the ESRL/GSD study in that they further hint at broader benefits over larger regions when the observations are added to the full current observing system.

Overall, the results from the two studies are supportive of the potential for positive forecast impact of GH observations for high-impact weather events other than tropical storms, particularly in the event of a gap in polar satellite observations. While the forecast improvements are smaller in magnitude than for the hurricane studies and typically lack statistical significance, the GSD analyses are just for a single storm system. The impacts of this storm from the SHOUT observational window were not as severe as might be desired for an analysis of this type. Analysis of more cases is clearly required to establish conclusive results for such winter-time sampling.

3 OBSERVING SYSTEM SIMULATION EXPERIMENT RESULTS

To complement and extend the data denial studies employing real observations, the SHOUT

project also supported a suite of Observing System Simulation Experiments (OSSEs) based on simulated observations. Simulated measurements with appropriate accuracy and resolution characteristics are sampled from an idealized, but realistic, truth state created by a "nature" run of a high-resolution numerical model. These simulated measurements are then assimilated into a different (ideally the current forecast system) model to evaluate their impact on helping that model simulate the desired truth state. The studies enable evaluation of a larger number of observation types and sampling strategies than fiscally feasible from actual airborne campaigns.

As for the data denial studies, the OSSE's focused on the impact of UAS-based observations on forecasts of hurricanes and high-impact Pacific winter storms. The hurricane studies were conducted using both regional and global scale models while the Pacific analyses were conducted with global models. The activities were conducted under direct support from SHOUT and in coordination with broader NOAA OSSE activities coordinated by Robert Atlas at AOML.

Results from the OSSEs were generally consistent with the evaluations of true observations and continue to support the potential positive value of UAS-based observations for forecasts of high impact weather events. In particular, the studies help illustrate the potential impact of different sampling strategies and the relative impact of different observation types.

3.1 Regional Hurricane OSSE Results

A regional hurricane OSSE capability established at AOML/HRD supported several analyses of relevance to SHOUT. Studies funded by SHOUT were led by Brittany Dahl while earlier analyses were conducted under the leadership of Robert Atlas.

3.1.1 Preliminary Studies

Two initial hurricane OSSE investigations, performed under the leadership of Robert Atlas, generally highlighted the potential for UAS observations to have a significant positive impact on hurricane forecasts. These studies utilized a nature run derived by embedding the high resolution (1 km) Advanced Research Weather Research and Forecasting (WRF ARW) model within the global ECMWF T511 (~45 km) resolution nature run. Data corresponding to three hurricanes observed in the global nature run (described in more detail in section 3.2) provided boundary conditions for the WRF ARW model. Forecast simulations within the regional tropical cyclone OSSE system were then performed using NOAA's HWRF model.

Using this system, experiments were conducted comparing forecasts assimilating standard observations other than those from UAS (the control case) with those additionally assimilating UAS dropsonde type observations. Assimilation of the data was possible both on the global and regional scales. Illustrative results were provided for the impact of assimilating dropsondes on the global scale and its subsequent effect on the initial and boundary conditions for the regional HWRF simulation. The impact, reported by Dr. Atlas, was primarily positive, indicating the potential for large-scale UAS observations to improve hurricane forecasts. Results obtained for first Atlantic basin hurricane appearing in the ECMWF nature run are summarized in Figure 3.1. The global assimilation of the UAS-based dropsonde data dramatically reduces the track error for

the hurricane, particularly in the first 2-3 days of the forecast, and results in a small but significant improvement in the intensity forecast as reflected by the minimum sea level pressure and maximum wind speed. The forecast underestimates the intensity at the start of the period, but becomes better as the lead time increases.

A second experiment provided further support for the potential positive impact of UAS observations on hurricane intensity forecasts. A common issue in regional hurricane modeling is that strong hurricanes are often forecast to initially "spin-down" in the early period of forecast runs. This impacts the short-term evolution of the vortex and hence potentially limits the predictability of intensity. Dr. Atlas reported the initial results of a study, in an OSSE environment, investigating whether there is a necessary minimum complement of observations that could eliminate this spin-down. The study employed the HWRF model with idealized, perfect observations both interpolated directly from the nature run and assimilated using the GSI system. The nature run was the same as described above (see first paragraph of Section 3.1.1).

Results are shown in Figure 3.2. The left panels shows the intensification of a strong hurricane from the WRF nature run. The center panels show the initialization and subsequent 6- hour HWRF forecast resulting from interpolation of perfect initial conditions from the nature run. With the near-perfect initial conditions, no spin-down occurs. The right panels show the corresponding results using the GSI system. While the representation of the hurricane is somewhat weaker in this case, the hurricane still does not spin down. The results support the hypothesis that the spin-down may be a result of inadequate representation of the vortex in forecast initialization. Additional work was planned to determine the minimum observational data needed to routinely eliminate the spin-down effect.

The findings demonstrate the value of providing a complete, accurate initial representation of strong vortices in hurricane forecasting. While multiple components of an observing system could contribute to supplying the required observations, UAS could potentially play an important role, providing high resolution observations throughout the entire atmospheric column with high spatial and temporal sampling density.

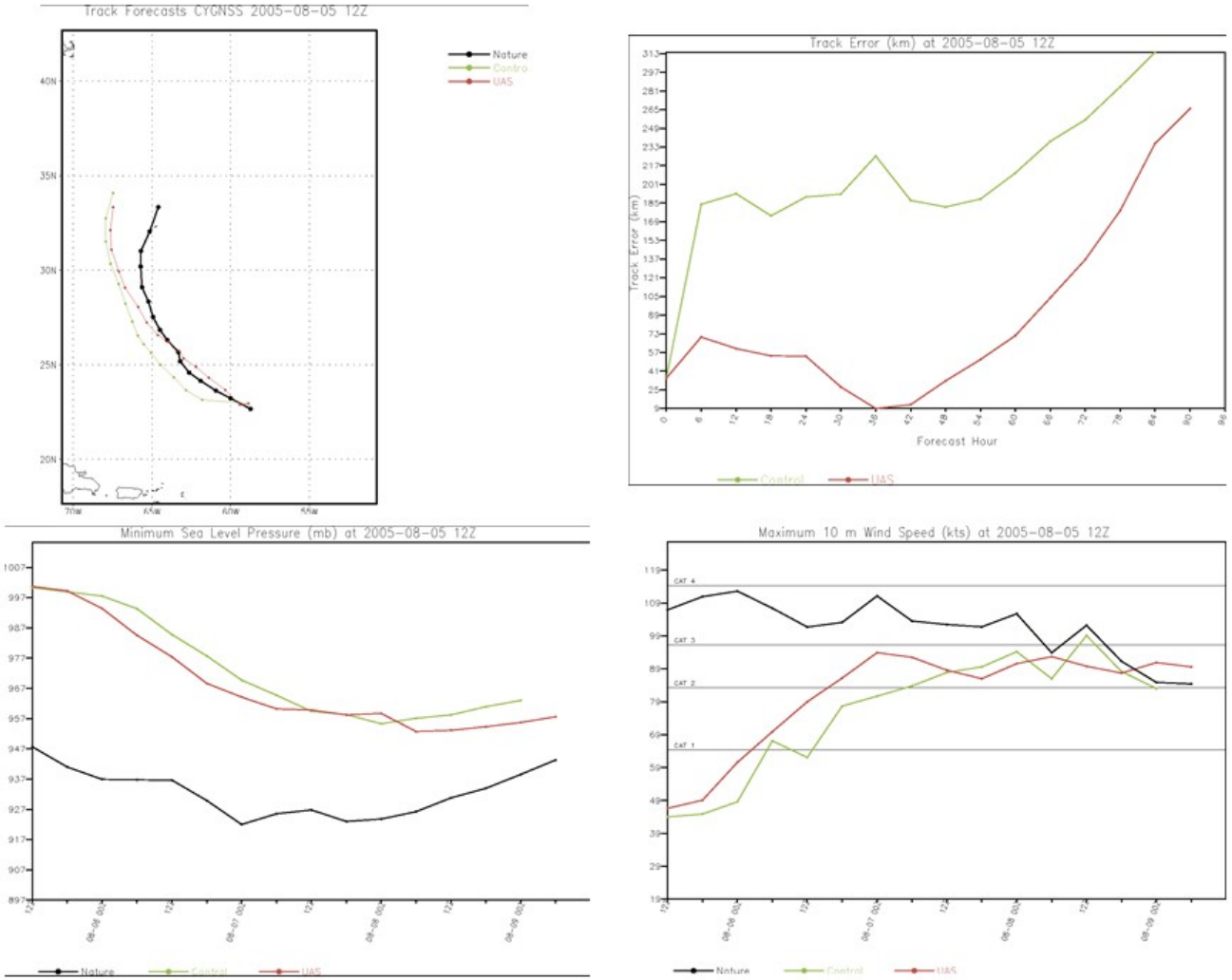

Figure 3.1. Results demonstrating the potential impact of global assimilation of UAS-based dropsonde observations on a hurricane forecast with the HWRF model performed within the regional tropical cyclone OSSE system. The case is for the first Atlantic basin hurricane observed in the ECMWF global nature run. The upper left panel compares the track forecasts for the control run assimilating standard observations only (Control, green) and assimilation of UAS-based dropsondes (UAS, red) with the best track from the nature run (black). The upper right panel displays the corresponding errors. The bottom panels demonstrate the impact on intensity forecasts as reflected by the minimum sea level pressure (left) and maximum wind speed (right). Graphics provided by Robert Atlas.

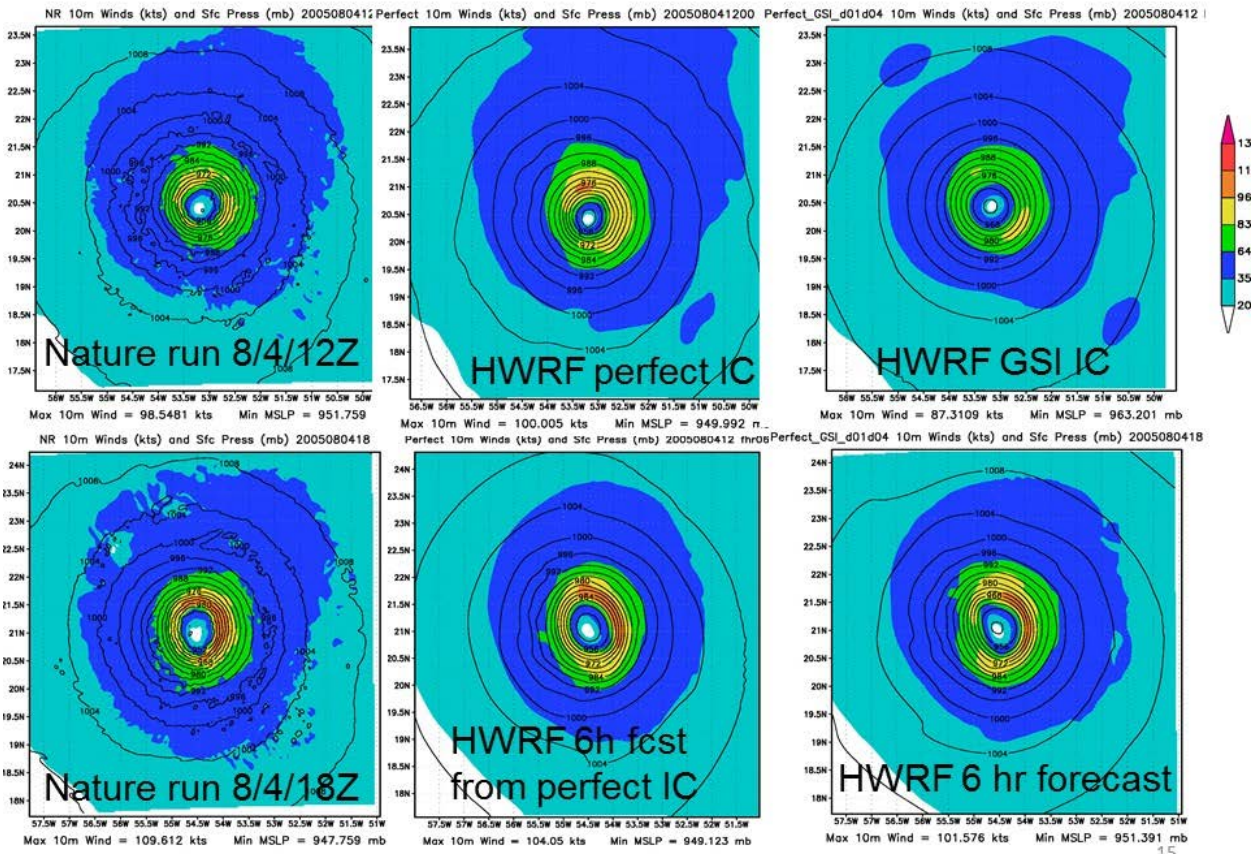

Figure 3.2. Evolution of the simulated hurricane wind speeds over a 6-hour period for the WRF nature run (left), HWRF with near perfect initial conditions (IC; middle), and HWRF with initial conditions from the GSI system (right). Graphics provided courtesy Robert Atlas.

3.1.2 SHOUT Supported Work

Hurricane OSSE work performed at AOML/HRD directly under SHOUT support employed the HWRF model with the HEDAS data assimilation system with sampling from the Nolan et al. (2013) nature run. The nature run includes a hurricane which intensifies from Category 1 to Category 3 strength over a short period beginning near 2 August 2005. The track and intensity of the storm over its lifetime as taken from the nature run are shown in Figure 3.3. Analysis of this system focused on a period from 2-5 August as highlighted in the figure.

The studies focused on evaluating what areas of the storm should be targeted to achieve the greatest forecast impact from GH observations and if sampling at different times with respect to the diurnal cycle had any preferential impact. In particular, the studies examined whether sampling closer to the center of the storm or further away from the inner core was most beneficial.

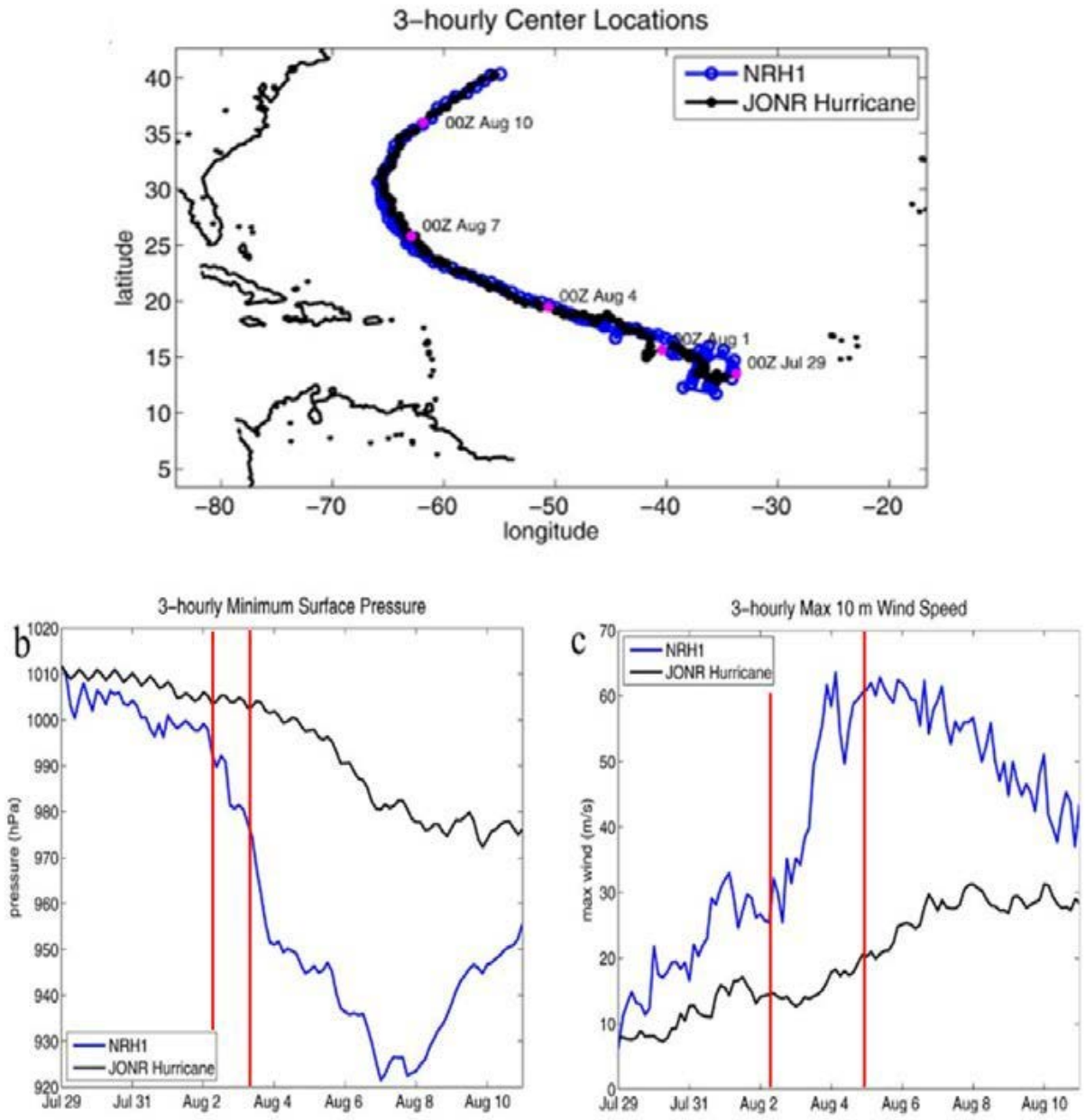

Figure 3.3. Characteristics of the nature run hurricane sampled in the OSSE analysis. The relevant traces are those in blue reflecting the actual nature run hurricane (NRH1). The black traces represent results from the joint OSSE nature run (JONR) from which the hurricane nature run was derived. The track is shown above while the corresponding evolution of the minimum sea level pressure and maximum wind speed are shown below. The primary analysis period ranging from 12 UTC on August 2 to 00 UTC on 5 August is indicated by the red vertical lines in the wind speed panel. Graphics provided by Brittany Dahl and Altug Aksoy.

Initial comparisons of the impact of sampling location were based on different simulated dropsonde locations along a common "rotated butterfly" pattern as shown in Figure 3.4. In the original pattern, 63 simulated dropsondes are spaced evenly throughout the pattern. In the supplemental cases, 28 additional dropsondes are inserted evenly or with enhanced sampling

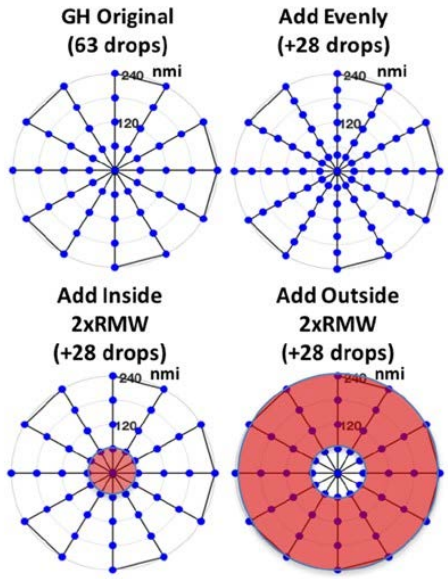

Figure 3.4. Global Hawk (GH) sampling patterns evaluated in the OSSE study. The top left panel shows the baseline "rotated butterfly pattern with 63 simulated dropsonde locations. The remaining panels show different modifications where 28 more dropsondes are added. In the top right, they are spaced evenly between the baseline locations. In the bottom left, they are all added near the storm center, within twice the radius of maximum winds (RMW). In the bottom right, they are all added further away from the center, outside twice the radius of maximum winds. Graphics provided by Brittany Dahl and Altug Aksoy.

either inside or outside twice the radius of maximum wind.

Overall results were obtained averaging the results of continuous cycling of eleven model runs at successive times from 12 UTC on 2 August to 00 UTC on 5 August and are summarized in Figure 3.5 relative to a control run with assimilation of standard observations only. The impact of the GH observations on the initial model analysis is observed to be positive for the storm position (track) and MSLP but limited for intensity as reflected by the maximum wind speed. For the initial position, unsurprisingly, the largest benefit is obtained for more observations near the center of the storm. With respect to MSLP, adding more observations including more sampling near the storm center improves the analysis. For forecasts, the impact of the observations on the track is short lived, but large and persistent for the MSLP and storm intensity. The positive impact of the observations on MSLP and maximum wind speed are consistently largest when more observations are added inside twice the radius of maximum winds. Potential improvements in excess of 40% are observed at forecast lead times beyond 12 hours. Separate analysis of individual forecasts at different stages of the storm (not shown) were consistent in reflecting positive impacts in forecasts of MSLP and intensity with the maximum benefit achieved with the addition of enhanced sampling near the storm center.

These OSSE results generally agree with the OSE studies described in section 2.1.3 in showing positive impact of GH dropsonde observations within HWRF-HEDAS forecasts with greatest benefit resulting from enhanced sampling near the storm center as opposed to at greater distances extending into the surrounding environment. The results have significance in that the HEDAS system is presently better suited to explore the benefit of inner core observations than within the operational HWRF system where observations within ~150 km of the center are excluded.

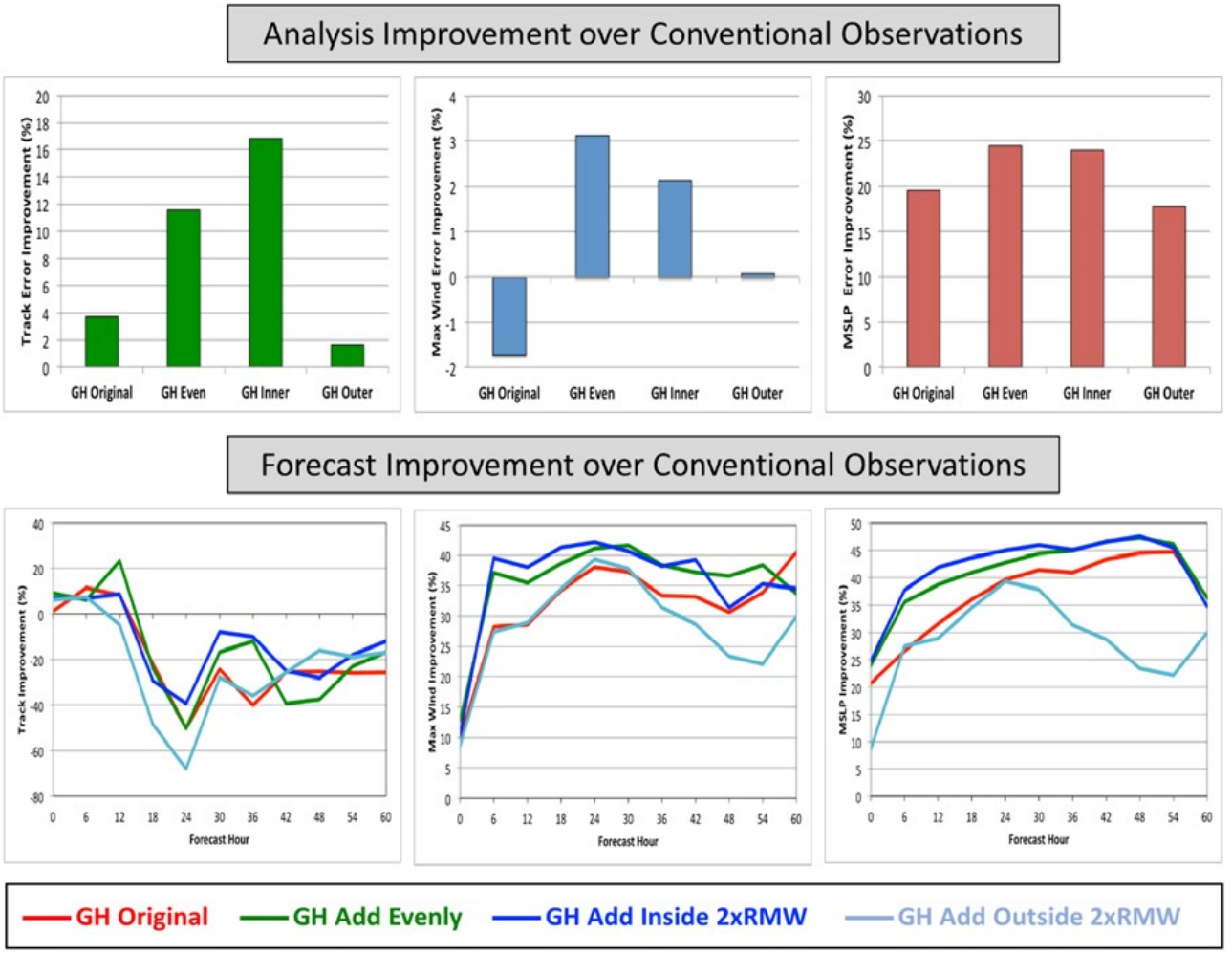

Figure 3.5. Summary of the impact of adding the simulated Global Hawk (GH) dropsonde measurements relative to a control run assimilating only conventional observations. The impact on the initial analysis is shown in the upper panels while the forecast impact is shown below. The left most panels are for the track error, the middle panels are for the maximum wind speed, and the right panels are for minimum sea level pressure (MSLP). The various different cases are as illustrated in Fig. 3.4 and described in the text. Graphics provided by Brittany Dahl and Altug Aksoy.

Comparison of independent model runs with sampling over different six-hour periods also provided initial insight into whether sampling at different phases of the tropical cyclone diurnal cycle affects the impact of the observations. Tropical cyclones are observed to exhibit a diurnal cycle (Dunion et al. 2014) where the cloud field regularly propagates radially outward from the storm. A baseline model run was defined for 12 UTC on 4 August corresponding to sampling at early morning and sunrise. Results were compared for runs six and twelve hours earlier and later using identical sampling based on the basic rotated butterfly pattern illustrated above. The results reflected changes in the impact of the simulated GH observations on both the initial analysis and forecast. The relative changes to the improvements in forecast error are shown in Figure 3.6. Over the first 12-18 hours, some improvement was obtained with sampling twelve hours earlier, while at longer forecast lead times, the most consistent benefit appeared to occur for sampling twelve hours later. Both of these correspond to sampling around 00 UTC in the early evening and sunset, suggesting the potential for benefit from ensuring sampling during this period of the day.

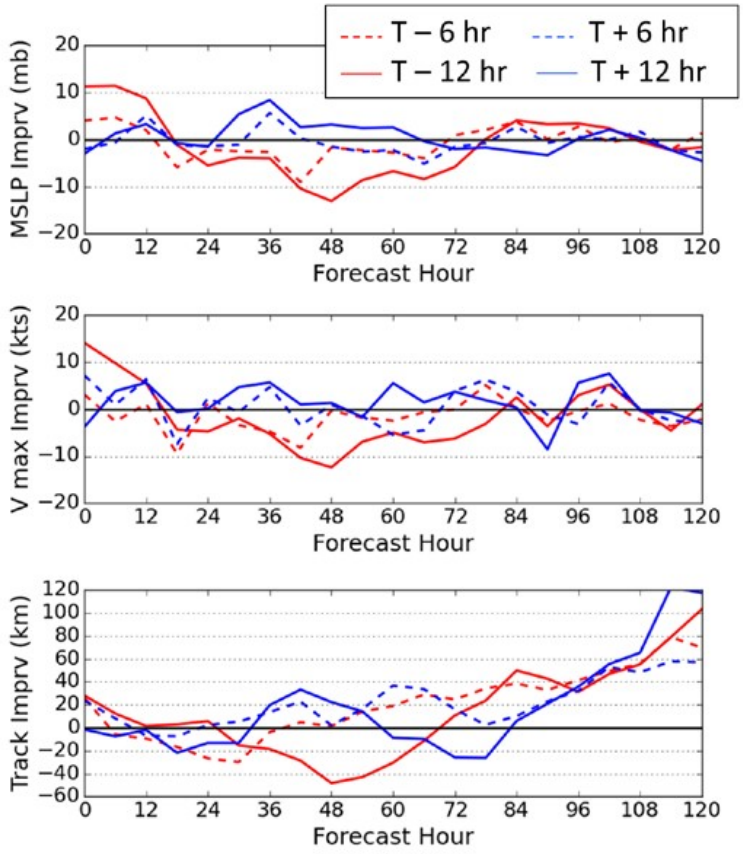

Figure 3.6. Summary of the impact of shifting the sampling time of the OSSE model run initially performed at 12 UTC on 4 August. The traces show the change in forecast accuracy relative to the initial forecast for periods offset by six-hour intervals as shown in the legend. The top panel is for minimum sea level pressure (MSLP), the middle for the maximum wind speed (V max), and the bottom for the track error. Graphics provided by Brittany Dahl and Altug Aksoy.

It is important to emphasize that one key advantage of the GH is that it is able to collect continuous observations throughout multiple phases of the diurnal cycle. The results, however, can help guide launch and recovery times of the aircraft to maximize overall sampling benefit.

3.2 Global OSSE Results for Hurricanes

A global OSSE system was initially established at ESRL prior to the start of the SHOUT program (Privé et al. 2013) and, over time, has been employed in multiple configurations to help evaluate the potential impact of high-altitude UAS observations on hurricane forecasts. A portion of the initial development activities was funded through the NOAA UAS program. The system, at the time, represented a significant improvement over previous capabilities through both the length and quality of the nature run, and the availability of more sophisticated synthetic observations. The nature run consisted of a 13-month free run of the ECMWF operational forecast model version c31r1 at T511 (~45 km) resolution. Synthetic observations available included conventional observations such as rawinsondes, dropsondes, and aircraft in situ measurements, and radiance observations from AIRS, AMSU-A, AMSU-B, HIRS-2, HIRS-3, MSU, and GOES. Calibration of the OSSE system was performed to enable attainment of more realistic results.

Initial hurricane results were obtained using the GFS forecast model with the GSI data assimilation package (Kleist et al. 2009) operational at NCEP in 2007. Using this system, the ESRL group conducted a preliminary study evaluating the potential for dropsonde observations from a high-altitude, long-endurance UAS to improve hurricane track forecasts. Though limited in scope, the results of the study did demonstrate the potential for improvement of the track forecasts. Within the nature run there were twelve hurricanes in the Atlantic basin, four of which made landfall in North America. While the cyclones are present, the inner structure of the storms was not well represented in the global model and thus the study focused on track rather than intensity forecasts. The preliminary investigation considered the impact on two forecasts of the first Atlantic basin storm within the nature run. Control forecasts generated using the standard set of available traditional observations showed substantial room for further improvement. The track of the selected storm and errors in the control forecasts are illustrated in Figure 3.7. The studies primarily considered circumnavigational flights where the UAS repeatedly sampled the region of the storm.

In the circumnavigation tests, the simulated UAS was flown in concentric circles around the moving center of the hurricane at various radii. The various configurations are summarized in Figure 3.8 and Table 3.1. The trajectories listed with repeated letters in Table 3.1 correspond to flights with the same number of repetitions of the patterns in Figure 3.8 to roughly match the endurance capabilities of the GH aircraft. In each trajectory, a simulated dropsonde was released every 105 km along the flight path. In some cases, the total number of sondes exceeds current capabilities of the GH dropsonde systems.

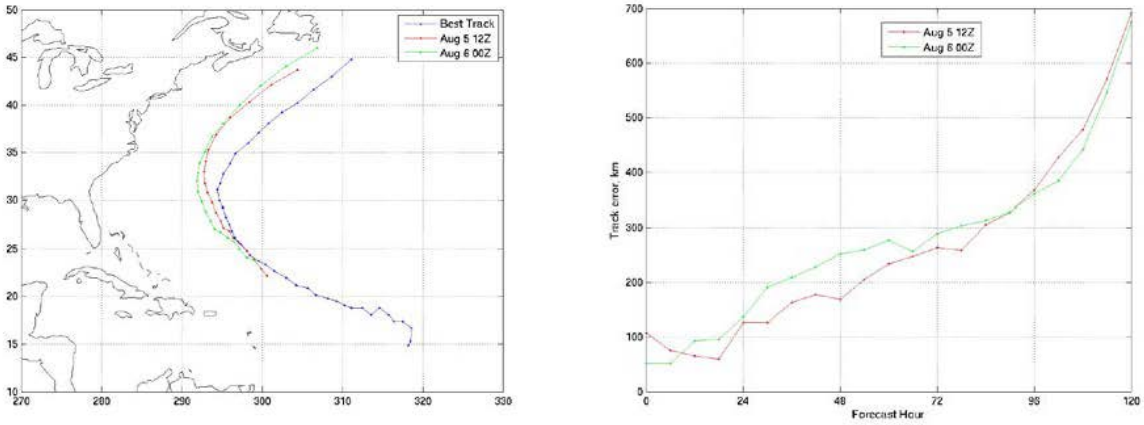

Figure 3.7. Control track forecasts of the selected storm (left) and illustration of the track errors (right). The best track from the nature run is shown in black while the control forecasts for two different starting times are shown in red and green. Graphics provided courtesy of Nikki Privé and Yuanfu Xie.

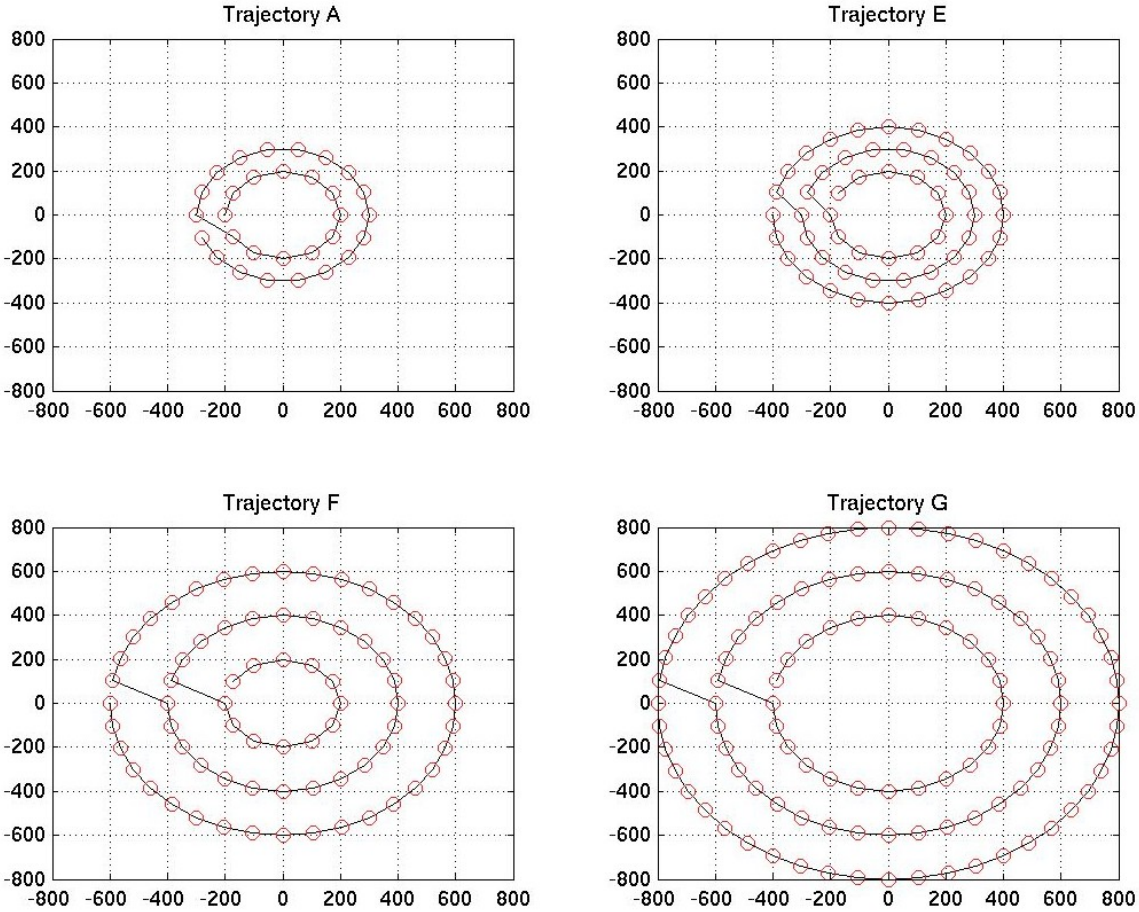

Figure 3.8. Illustrations of the trajectories tested in the initial ESRL global OSSE study. Simulated dropsondes, indicated by the red circles are spaced 105 km apart. The radii for the trajectories are as follows: A – 200 and 300 km; E – 200, 300, and 400 km; F – 200, 400, and 600 km; G – 400, 600, and 800 km from the storm center. Graphics provided courtesy of Nikki Privé and Yuanfu Xie.

Table 3.1. Details on flight scenarios evaluated for the initial ESRL global OSSE study. Results provided courtesy of Nikki Privé and Yuanfu Xie.

Trajectory	First Observation (Zulu Time/Date)	Flight Time (h)	Number of Sondes
Trajectory AA	06Z 5 Aug	11	60
Trajectory AAA	06Z 5 Aug	16	90
Trajectory EE	03Z 5 Aug	19	108
Trajectory F	06Z 5 Aug	13	72
Trajectory FF	12Z 4 Aug	26	144
Trajectory G	06Z 5 Aug	18	108

For these experiments, the forecast hurricane tracks show slight improvement in the along-track error, but little improvement in the cross-track error. The forecast tracks for initialization at 12Z on 5 August are shown in Figure 3.9 and the corresponding track forecast errors are summarized in Table 3.2. The main effect of the UAS observations is to slow the too-rapid northward propagation of the storm in the first few days of the forecast. As seen in Table 3.2, the track error was reduced at all forecast times. The impact on the forecast fields was primarily seen within a 1000-km radius of the storm center, consistent with the observational region. The greatest improvements were achieved for the trajectories with the larger radii. While the focus was on track forecasts, assimilation of the dropsonde observations did also result in reductions in the error in the mean sea level pressure of the hurricane environment.

Additional refinements were made to the global OSSE system at ESRL prior to the start of SHOUT under funding from the UAS Program to support additional tropical cyclone studies. A primary update was for the use of the global GSI Hybrid Variational Ensemble Kalman Filter data assimilation system (GDAS). The updated studies looked at the impact of the newer assimilation system on hurricane track forecasts and employed an ensemble transform sensitivity method to identify data sensitive regions for the hurricanes.

Two hurricane cases from the original nature run were tested using the revised assimilation system. The first was the same event explored in the preliminary study described above and the second was the second Atlantic basin hurricane of August. Use of the hybrid GDAS with just the standard observations produced significantly improved track predictions for the first event relative to the old OSSEs leaving little room for the simulated UAS dropsonde data to provide further improvements (results from Yuanfu Xie, not shown). Overall, however, the results reflected a notable improvement in global modeling capabilities enabling more detailed future testing with potentially higher resolution.

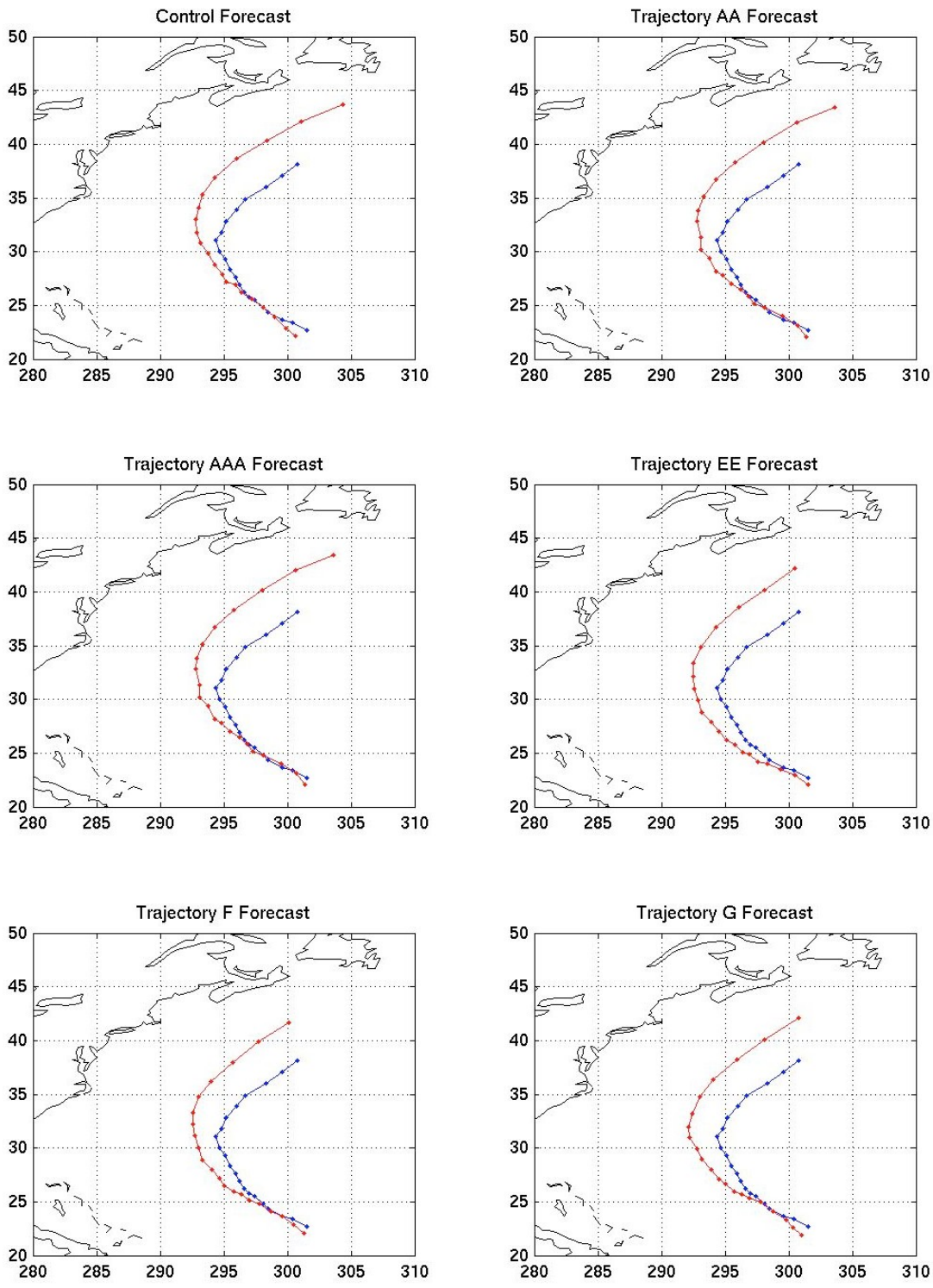

Figure 3.9. Comparison of forecast tracks (red) with the best track derived from the nature run (black) for the hurricane forecast initialized at 12 Z on 5 August. Graphics provided courtesy of Nikki Privé and Yuanfu Xie.

Table 3.2. Track forecast errors for the various trajectories and control as shown in Figure 3.3. All values are given in km. Results provided courtesy of Nikki Privé and Yuanfu Xie.

	Analysis	24 h	48 h	72 h	96 h	120 h
Control	108	127	170	264	367	692
Trajectory AA	67	92	171	210	347	635
Trajectory AAA	67	92	171	210	347	635
Trajectory EE	67	84	134	174	288	456
Trajectory F	70	30	127	164	293	405
Trajectory FF	24	52	169	186	285	571
Trajectory G	103	38	121	183	293	445

Use of the ensemble transform sensitivity method enabled a more direct identification of individual regions with highest forecast sensitivity than the preliminary methods and tests employing sampling of these regions were performed under the leadership of Yuanfu Xie. A summary of the experiment results for the second Atlantic basin hurricane is shown in Figure 3.10. The upper left panel shows the results of the sensitivity analysis where the shaded regions highlight the regions with greatest identified forecast sensitivity. Based on these an experimental sampling pattern with dropsonde releases was developed as shown in the upper right panel. The impact of this sampling was compared with a control case with no additional assimilation and an assumed complete limiting sampling shown by the grid in the lower right panel. The results as shown in the lower left panel demonstrate that all forecasts are quite good over the first 1.5 days and then the maximal sampling does lead to an improved track forecast over the next 1.5 days. The limited targeted sampling employed in this case demonstrates only a small improvement over the control forecast at longer lead times.

Overall, while the early global hurricane OSSE results showed some positive impact of the assimilation of dropsonde observations within a global model, the improvement in the forecasts was generally small. Much greater observational impact was obtained for real dropsonde measurements in the current operational GFS modeling system as presented in Section 2.2.2. The differences likely highlight limitations in the global OSSE analyses given the coarser resolution of the nature run and the age of the assimilation methods employed. Perhaps the most significant result from these early studies was the suggestion that the greatest positive track impact came from sampling more distant from the storm center. Potential benefits of the OSSE approach for comparing the relative impact of different new sensor and observation types were not fully exploited for global hurricane studies. SHOUT-supported global OSSE work focused in more detail on analysis of midlatitude storm systems as described in the following section.

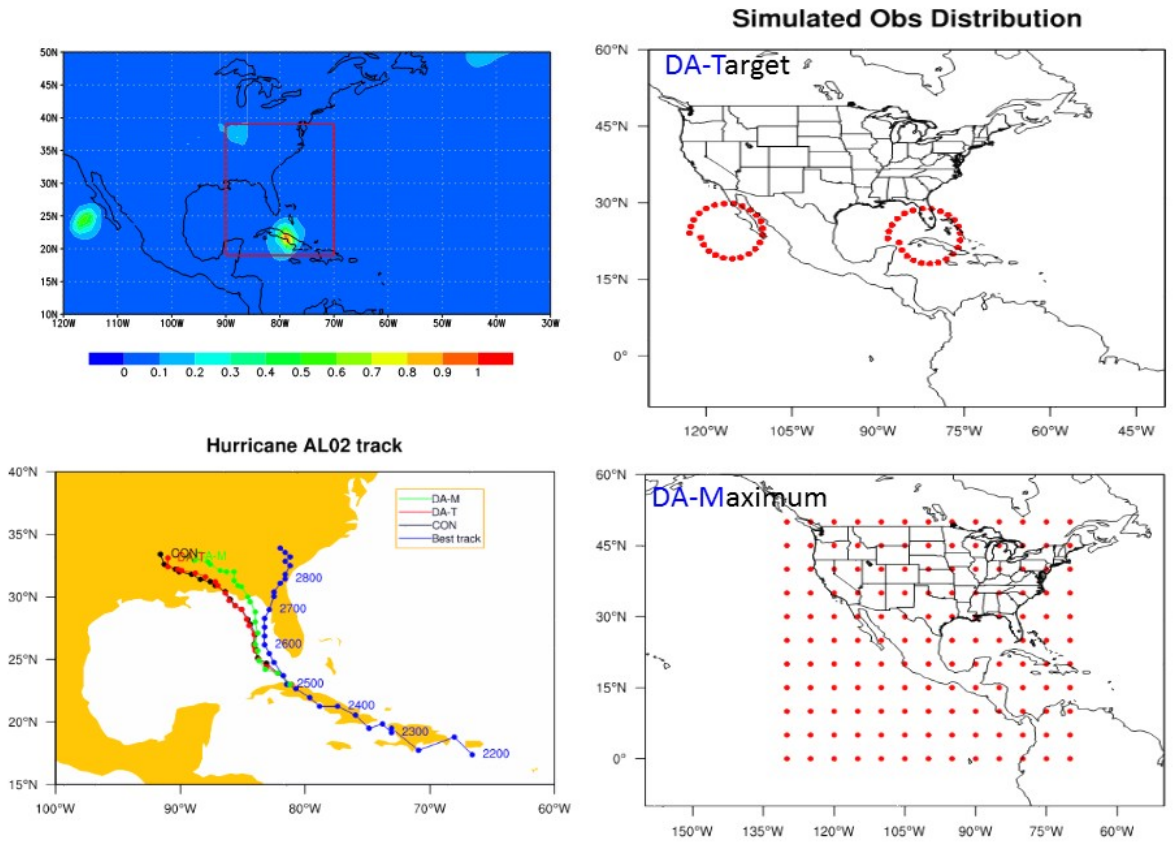

Figure 3.10. Example experiment employing the refined global OSSE system. The experiment is for the second Atlantic basin hurricane observed in the nature run and the forecasts were initialized at 00Z on 25 August 2005. The upper left panel shows shading of the identified sensitive regions. The right panels show the simulated distribution of dropsonde/radiosonde observations (Obs) for sampling of the sensitive regions (top) and a theoretical data assimilation (DA) maximum observation density case (bottom). The bottom left panel shows the forecast tracks with the simulated observations compared against a control case with no data assimilation and the best track derived from the nature run. Graphics courtesy Yuanfu Xie.

3.3 Global OSSE Results for Midlatitude Storms

More recently, the GOSA team at ESRL/GSD led by Lidia Cucurull applied the global OSSE system to studies related to flights targeting high-impact weather events other than tropical cyclones. The goals focused on exploring different potential sampling strategies and evaluating the impact of different measurement types to help guide future payload selection.

These recent global OSSE studies used the same ECMWF T511 nature run and the NCEP GFS model with updated GDAS operating at the lower T382 resolution. A higher resolution nature run with synthetic observations was recently completed, but not in time to be employed in the studies. The work had three components with increasing complexity: 1) an ideal study assimilating perfect observations of temperature, wind, pressure, and moisture at every grid point over a large area; 2) sampling with perfect observations from an identified sensitive, but still large, area; and 3) sampling from actual flight paths with more realistic observations. All of

these elements were considered both with and without satellite observations to explicitly address satellite gap mitigation.

Work within the first element, while unrealistic from the point of view of true aircraft sampling, enabled comparison of the relative impacts of different types and combinations of observations. The impact of idealized dropsonde-type observations collected throughout large spatial domains when added to the complete current observing system was analyzed for three winter storm systems present in the nature run. The storms were termed the “Jan 29,” “Jan 30,” and “Feb 25” storms based on data from 2006 within the nature run. The track of the storms, selected verification regions, and the idealized sampling domains are shown in Figure 3.11. The “Jan 29” storm impacted Alaska, the “Jan 30” storm tracked south along the West Coast affecting Oregon before turning east and tracking through the Southwest, and the “Feb 25” storm tracked more easterly across the US with impacts extending into the Midwest. The sampling domains were selected, in part, due to patterns of analysis errors in a control forecast, without additional data assimilation, relative to the nature run.

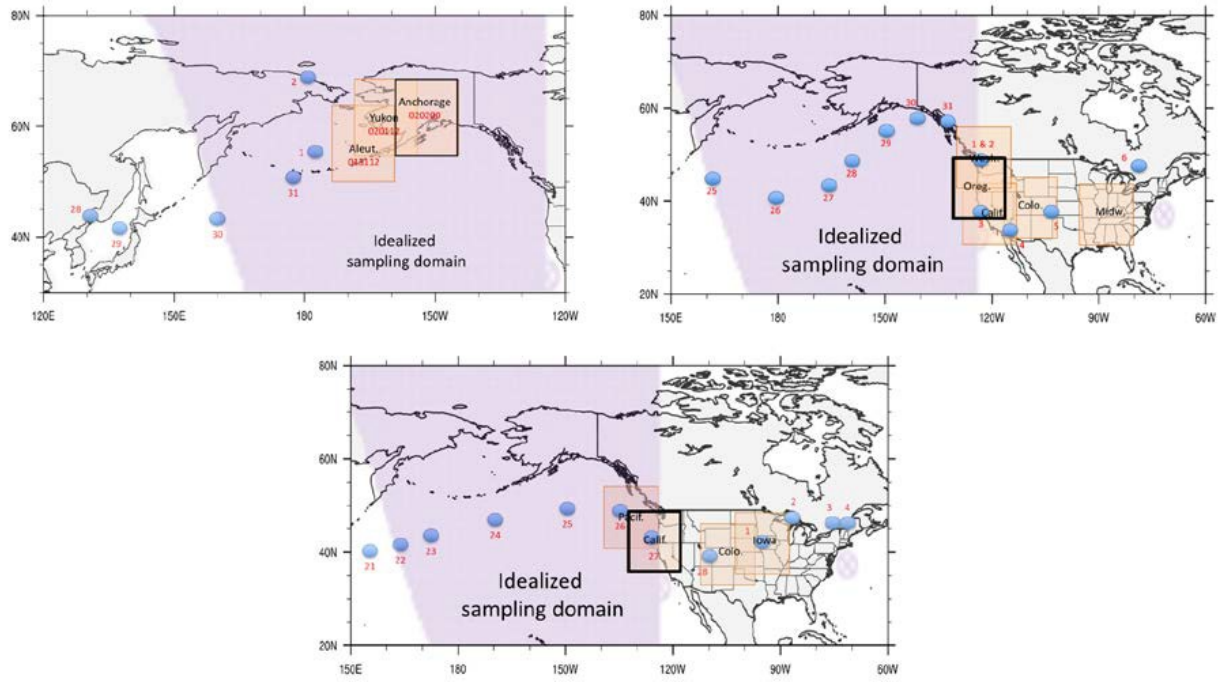

Figure 3.11. Illustration of the storm tracks (blue dots), verification regions (orange rectangles), and idealized sampling regions (purple shading) for three storms evaluated in the global winter OSSE work. The top left panel shows the 29 January (2006) storm, the top right the 30 January storm and the bottom panel shows the 25 February storm as taken from the nature run. Graphics taken from Peevey et al. 2017.

For each storm, experiments included assimilation of all possible variables, winds only, temperature only, and humidity only. The initial control for each case included assimilation of all conventional observations without the dropsondes. For each storm and case, a series of GFS model runs was performed every 12 hours, examining the impact of the observations for

forecast lead times from 1-7 days. The average change to the forecast error resulting from assimilation of the different observation types is shown in Figure 3.12. The errors are expressed in terms of a total energy metric and are plotted as a function of both forecast hour and verification date.

Not surprisingly, the greatest positive forecast impacts typically result from assimilation of all the observation types. The positive impact is consistently largest for the “Jan 30” storm, but is generally positive for each storm. For the individual observation types, the greatest positive impact generally results from either the wind or temperature data. Overall, the humidity data appears to have the least positive impact. The impact of the wind and temperature data is positive over nearly all lead times for both of the storm events. The energy error measure was selected for direct comparison with the results of Hamill et al. (2013) which studied the effectiveness of the previous operational winter storm reconnaissance program. Additional verification metrics such as 500-hPa height and sea level pressure were considered in addition to the total energy measure, and it was found that choice of the metric did not significantly change the conclusions on observation impact.

Experiments next examined the impact of sampling over more realistic domains, though still with perfect observations. The computed observation sensitivity and simulated flight tracks for the winter storms are illustrated in Figure 3.13. The impact of idealized observations was compared for sampling over the full idealized domain (Figure 3.11), all grid cells where the computed sensitivity exceeded a threshold of 0.5, and for realistic numbers of dropsondes deployed along the simulated flight track. The results, expressed as a change in forecast skill as a function of verification date, are shown in Figure 3.14. While the impact of fewer observations is clearly reduced at most times, the targeted sampling is still generally able to produce a forecast improvement over the current full observing system. The impact is typically small, with improvements of less than 5% for sampling from dropsondes alone along the flight track. Introduction of realistic measurement errors could further reduce the impact, but addition of continuous sampling along a swath under the aircraft using a remote sensor like HAMSR could potentially increase the impact towards the limit of full sampling of the sensitive regions.

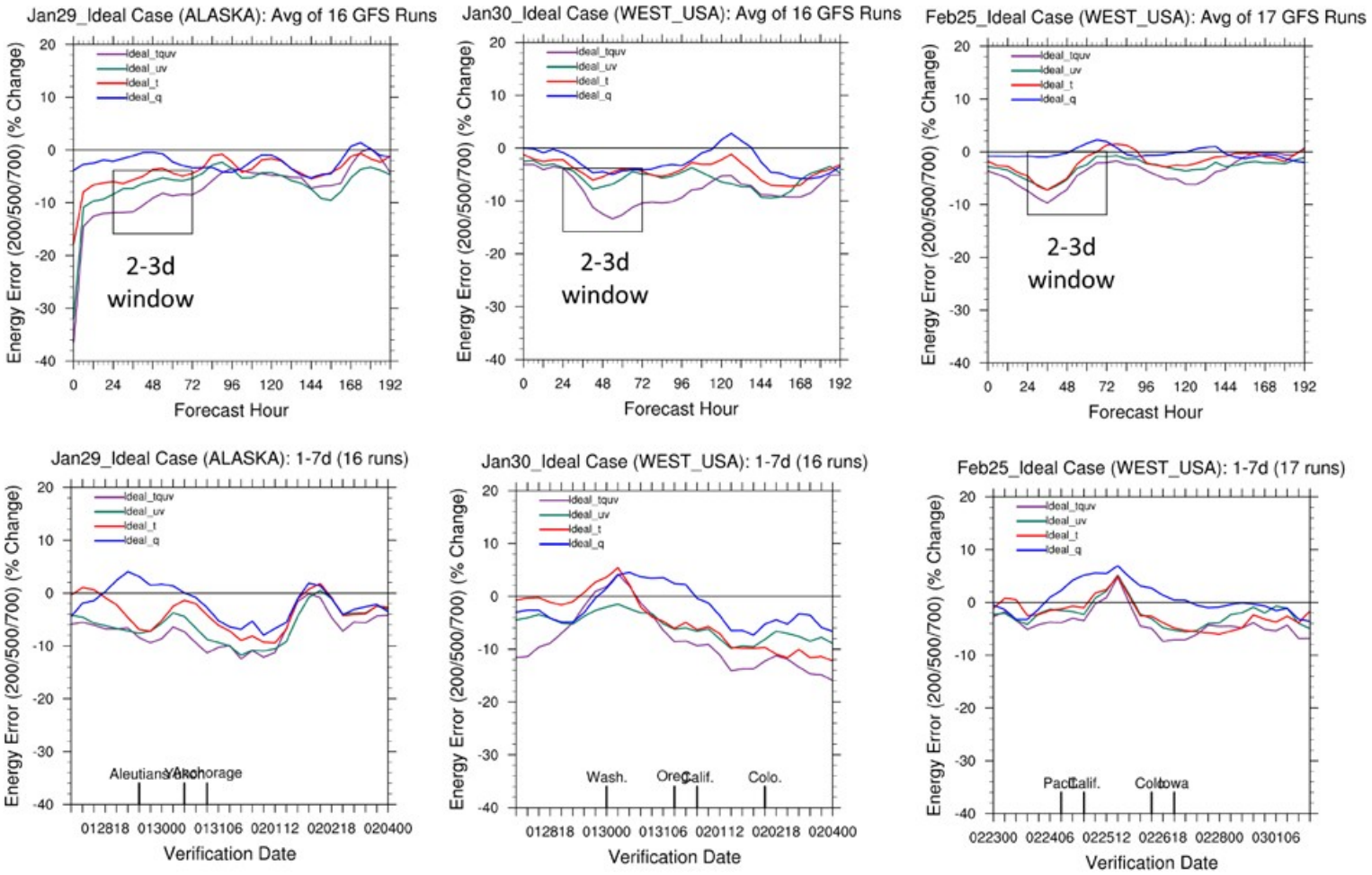

Figure 3.12. Comparison of forecast errors resulting from idealized sampling of different variable types for the three OSSE storms. Errors are displayed as a percent change in a total energy metric relative to the control case with assimilation of all standard observations. Results for the Jan 29 storm are shown on the left, the Jan 30 storm in the middle, and for the Feb 25 storm on the right. The upper panels show the errors evaluated within the verification regions (solid black rectangles in Figure 4.4) as a function of forecast hour, while the lower panels show the results as a function of verification date as the storms propagate downstream. Differences in sampled variables are indicated by the colored traces as denoted in the legend. Graphics provided by the GOSA Group and consistent with Peevey et al. (2017).

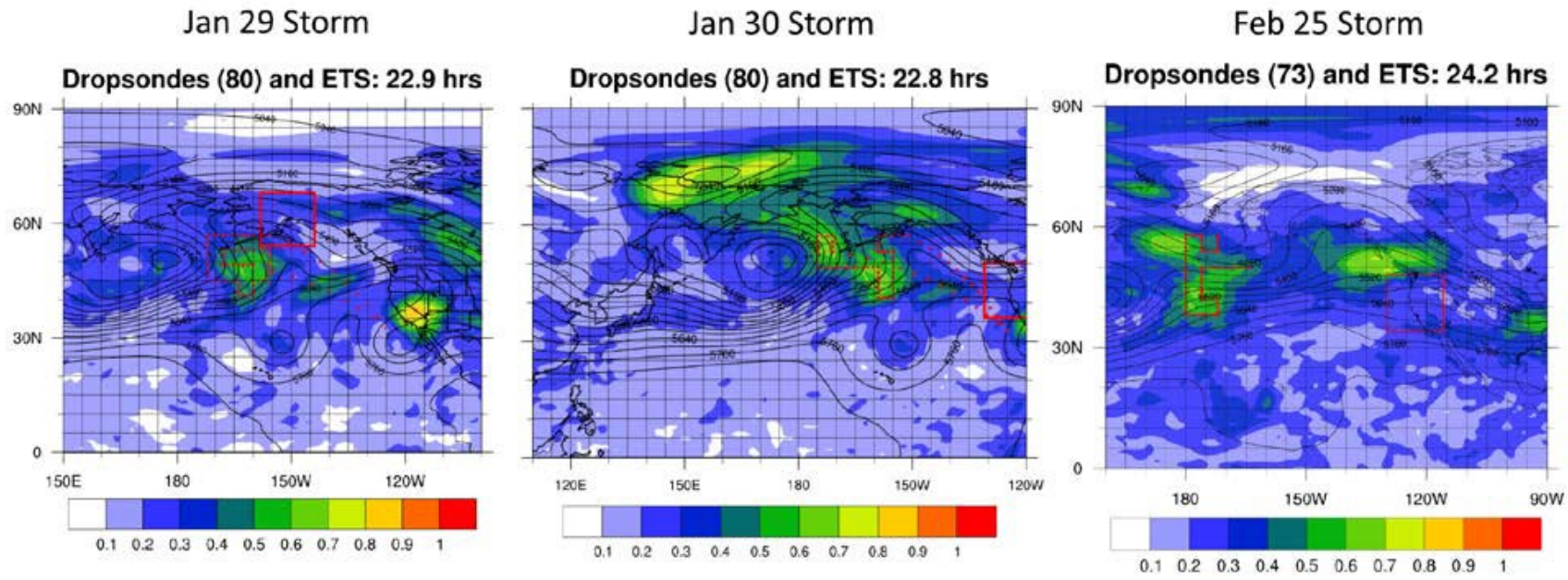

Figure 3.13. Computed forecast sensitivity and derived GH flight paths for the Jan 29 (left), Jan 30 (center) and Feb 25 (right) OSSE storms. The forecast sensitivity computed using the Ensemble Transform Sensitivity (ETS) methodology is shown by colored shading with warm colors indicating greatest sensitivity. The regions for which the forecasts were to be improved are shown with the red rectangles. The flight track (red dots indicating dropsonde locations) was computed by an automated procedure designed to optimize sampling of the sensitive areas under realistic GH operating constraints. Graphics provided by the GOSA Group as appearing in Peevey et al. (2017).

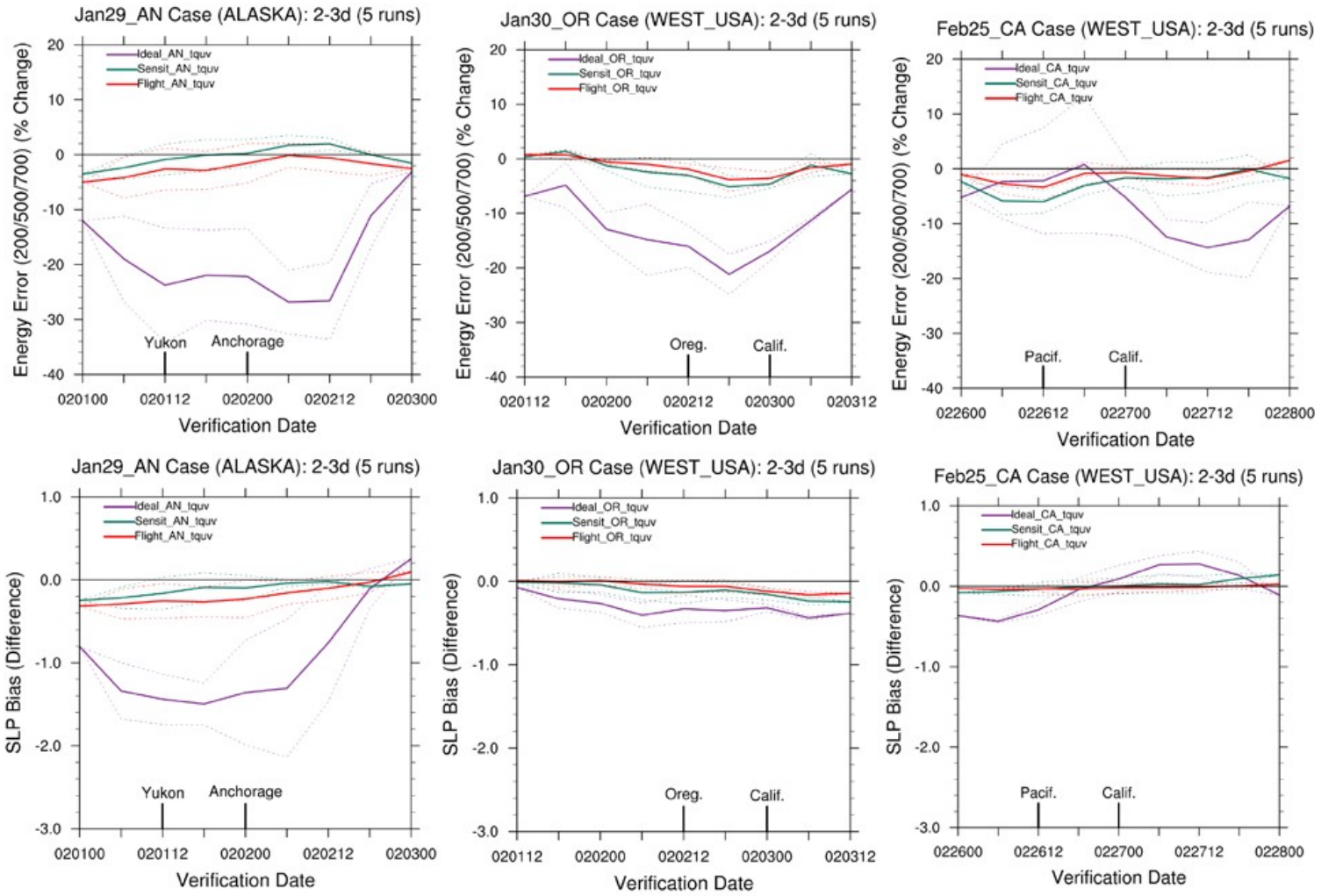

Figure 3.14. Comparison of the impact of idealized dropsonde observations sampled over different domains for the three OSSE storms. Results are shown for the percent change in the total energy error metric (top) and sea level pressure bias (bottom) computed over the verification domain and plotted as a function of verification date. The change is again computed relative to a control with assimilation of all standard observations. The different sampling regions are denoted with the different colored traces – purple: full idealized domain; green: all grid points with sensitivity above a threshold; red: sampling only at the drop locations along the GH flight track. Dotted lines reflect variability in the computed error change over the different model runs. Graphics provided by the GOSA Group consistent with Peevey et al. (2017).

To further explore how the selection of realistic GH flight tracks influences the amount of forecast improvement, results were compared for tracks emphasizing dropsonde-based sampling of different regions and features. In addition to the previous tracks that optimized sampling of the computed sensitive regions, tracks were developed to sample different specific meteorological features present during the flight period. For the “Jan 30” storm impacting Oregon, the features included a ridge present of the U.S. west coast, the jet exit region, a region of enhanced vorticity off the coast, and a region of elevated moisture content south of Alaska. For the “Jan 29” storm impacting Alaska, the features included a rapidly developing low pressure region, the jet exit region, and an atmospheric river. The impact of the sampling choice on the RMSE in modeled geopotential height at two levels is shown in Fig. 3.15 for the two storms. For the “Jan 30” storm, the sampling of the sensitive region identified with the Ensemble Transform Sensitivity (ETS) methodology provided the largest overall reduction in forecast errors. In contrast, on Jan 29, the targeting of the jet exit region resulted in greater forecast error reduction.

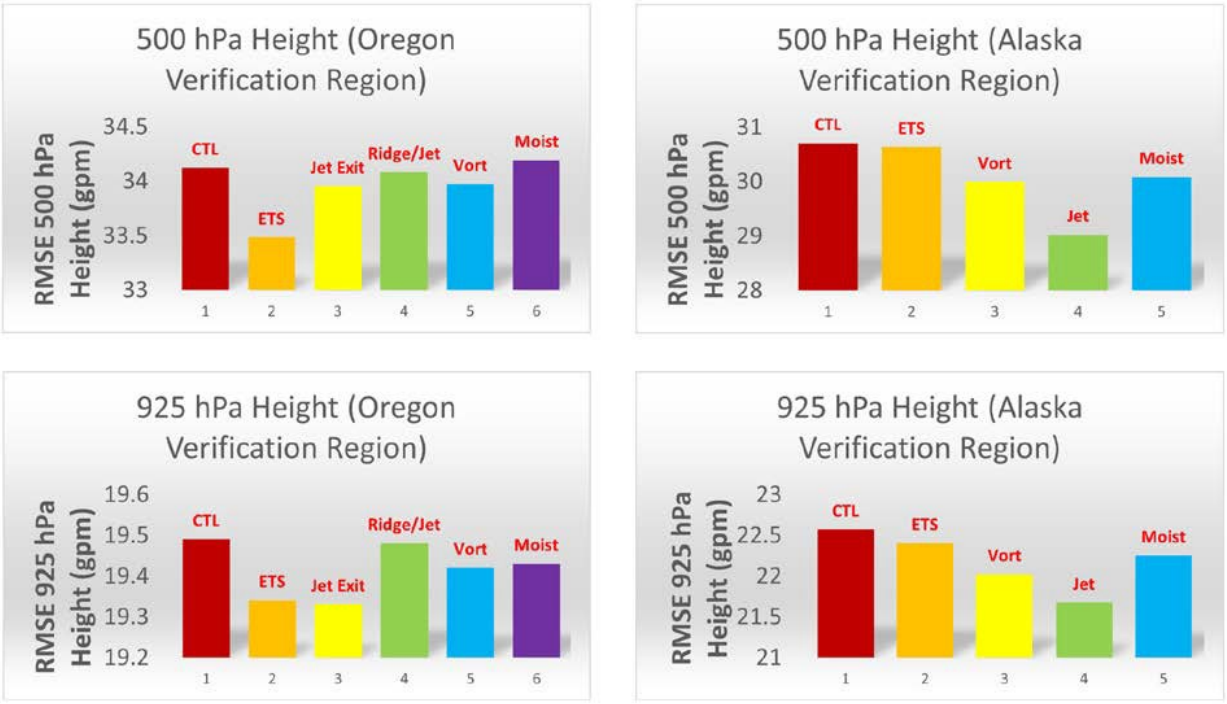

Figure 3.15. Summary of results illustrating how the selected target for sampling impacted the forecast root mean squared error (RMSE) in geopotential height relative to a control forecast (CTL) with assimilation of conventional observations only. The left panels are for the Jan 30 storm and the right panels are for the Jan 29 storm. The upper panels are for the 500-hPa height while the lower panels are for the 925-hPa height. The specific regions are as described in the text. Graphics provided by the GOSA Group.

The results, as a whole, support the potential to obtain forecast improvements of wintertime storms with realistic deployments of idealized dropsonde sensors. While targeted observing of midlatitude storms with a platform like the GH can lead to improved forecasts of high impact

weather events, the improvements are typically small when the observations are collected in addition to the current observing system. Of relevance to the SHOUT project objectives is the question as to whether the impact of the GH observations would be greater if there were an unexpected gap in satellite coverage.

To explicitly assess the potential impact of GH-type observations in the absence of key components of our current satellite observing system, several additional experiments were performed for one of the winter storm events considered previously. The experiments consisted of a control case with all current observations (as presented previously), a "gap" case with satellite observations withheld, and addition of idealized GH dropsonde-type observations to the gap case. Within the gap case, observations were withheld from the NASA Aqua, Defense Military Satellite Program (DMSP), NOAA 14-19, and Suomi NPP satellites. This is an extreme number of satellite observations, so the results should represent a limit of maximum possible impact from the GH observations as opposed to an expected, practical value. The "Feb 25" storm event was selected because it was a strong event and the idealized dropsonde observations had the least impact in the previous study relative to the full observing system.

Forecast sensitivity calculations were performed for both the control and gap cases based on a target impact region along the west coast, and GH flight tracks were designed to sample the regions of highest sensitivity as shown in Figure 3.16. The sensitive regions are generally similar in location, but the derived flight tracks do have differences. Multiple GFS runs were conducted over the lifetime of the storm and results comparing the average forecast error for the different experiments are shown in Figure 3.17. Removing the satellite observations results in a significant increase in the forecast error over the control case. Adding simulated dropsonde observations over the complete large idealized sampling domain (shown in Fig. 3.11) is able to largely compensate for the gap in satellite observations, but this domain is larger than can be sampled with a single GH. Adding the simulated dropsonde observations over only the identified sensitive regions or along the derived flight track are also seen to reduce the error relative to the gap case, but to a lesser degree. For the verification domain centered on California, removing the satellite data resulted in a forecast degradation of 2-6% for lead times between 2-3 days. Adding dropsondes along a realistic GH track provided up to a 4% improvement in the forecast relative to the gap case between 48-68 hours within the California domain but had negligible impact on global scale forecast skill.

The results, though just for one case, suggest that targeted observations from a platform like the GH can help mitigate the effect of potential gaps in satellite coverage for forecasts of high-impact winter weather over regions subject to the greatest impacts. The positive benefits, however, could largely be constrained to those limited regions. These results are broadly consistent with what has been observed in the global-scale OSE studies for midlatitude weather events as well.

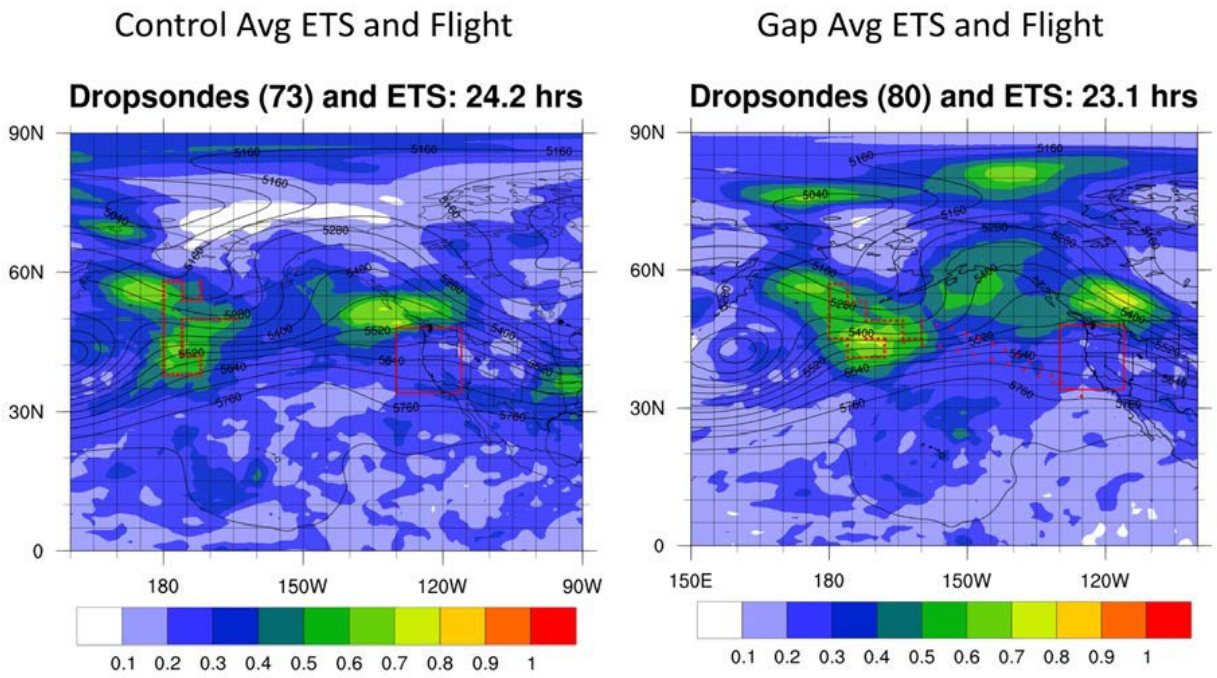

Figure 3.16. Comparison of the computed forecast sensitivity and derived GH flight paths for the Feb 25 storm depending on whether standard satellite observations are (left panel) or are not (right panel) also assimilated. The forecast sensitivity computed using the Ensemble Transform Sensitivity (ETS) methodology is shown by colored shading with warm colors indicating greatest sensitivity. The left panel is identical to the right panel in Figure 3.13. Graphics provided by the GOSA Group.

The SHOUT OSSE studies to date have not yielded significant guidance on the relative merit of specific individual sensor payloads. The results have largely focused on the forecast skill element using idealized observations. While the results above emphasize the importance of wind and temperature profile observations, the findings are not sensor specific. Studies related to radiance assimilation and sensor characteristics had to be delayed and were not completed during the SHOUT project duration. The general OSSE system, however, is now in place, and additional work is planned or proposed following the completion of SHOUT.

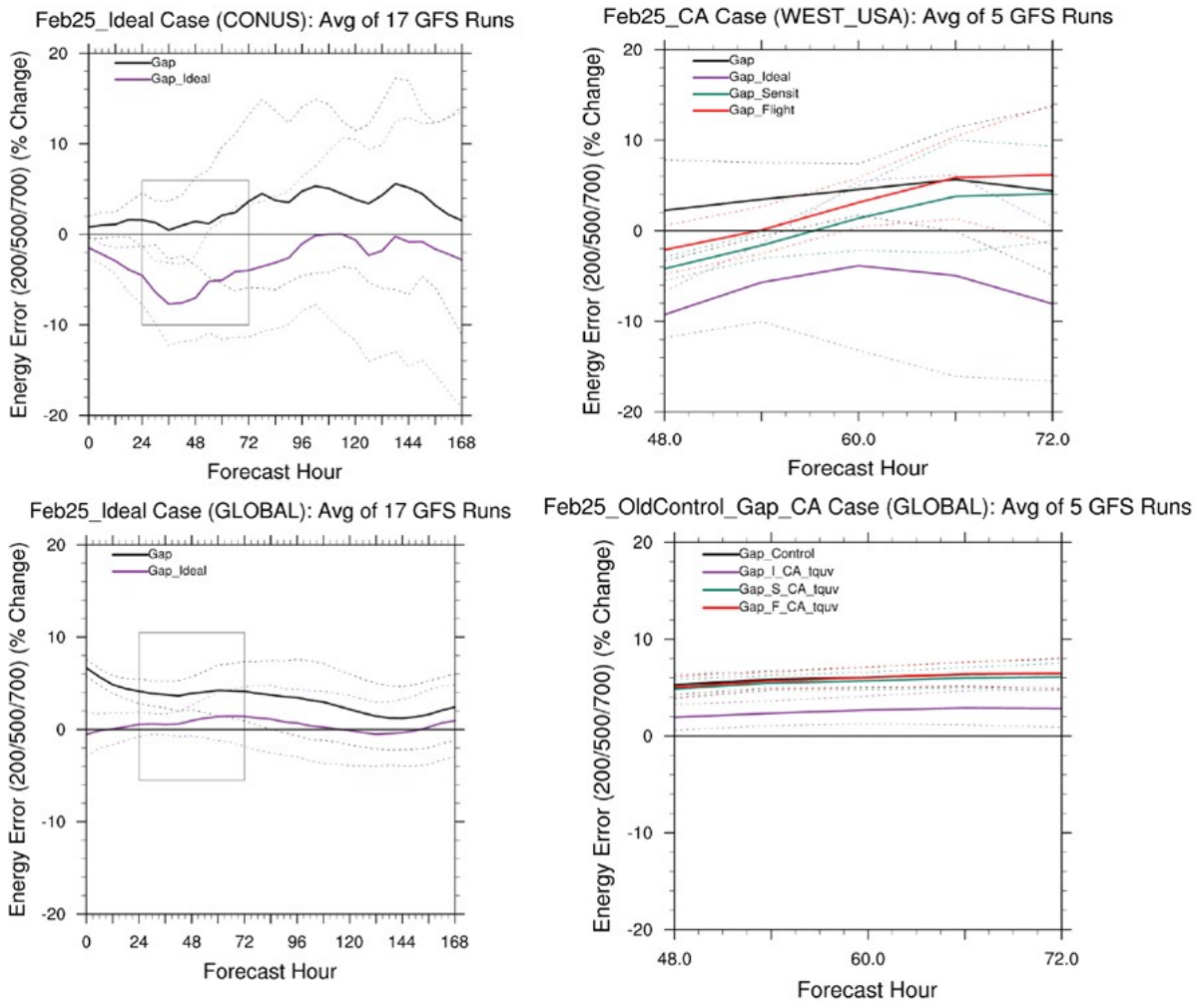

Figure 3.17. Comparison of the impact of idealized dropsonde observations sampled over different domains in the event of a gap in satellite observations. All results are displayed as the percent change in the total energy error metric relative to the control case including assimilation of all standard satellite observations. The black traces show the change just by excluding the satellite observations while the colored traces show the impact of adding the idealized dropsondes over the different domains as in Figure 3.16. Dotted lines again reflect variability across the different model runs. Results are shown for errors computed over different regions as indicated in the titles of the individual panels. Graphics provided by the GOSA Group.

4 CONCLUDING ASSESSMENT

Through SHOUT and collaborative work at NCEP/EMC, several diverse but complementary studies consistently demonstrate significant positive forecast benefits from including targeted observations from an unmanned aircraft like the Global Hawk during high-impact weather events. The results obtained at EMC with the current operational modeling system, in particular, are highly positive and argue strongly for the potential merit of the observations. Notable forecast improvements are also observed when existing elements of our satellite observing system are withheld, simulating the value of the GH data in the event of a possible

gap in polar satellite observations. The observed benefits span both regional and global models. Larger sample sizes would be desirable to increase statistical significance, but the results are highly encouraging and support the potential for forecast benefit from operational utilization of a platform like the GH. While an ultimate decision on utilization of the GH will incorporate budgetary considerations, the scientific value of the observations appears to be broadly supported.

Where presented, the impact of GH-type observations on the initial model analyses was uniformly large and positive. Significant improvements in storm structure were observed within both HWRF and global models. For the HWRF-HEDAS system, the improvements in the analyses were particularly large, exceeding 25%. This demonstrates that the observations can clearly improve the initial representation of the storms in the models and is one of the most direct tests of the value of the observations. The improvements were largely independent from the forecast system and associated complexities.

Impacts on forecasts were more varied depending on the modeling system, lead time, target, region, and variable, but, taken as a whole, the results demonstrate strong potential for achieving significant gains in forecast accuracy for high-impact events. An illustration of the broad potential for significant forecast accuracy improvements across research and operational modeling systems is presented in Table 4.1 which summarizes the results of the multiple studies on forecasted track accuracy for hurricanes and tropical storms at a 96-hr lead time. The potential for overall gains of 12% and gains in excess of 25% for individual systems within the current operational modeling systems is remarkable. The corresponding forecast intensity was also improved by up to 8% in the operational HWRF model at 96 hours and 14% at 72 hours. Intensity forecasts in the GFS model were not significantly improved, but that is not the model's primary focus.

Table 4.1. Global Hawk dropsonde impact on 96-hour tropical cyclone track forecasts (i.e., percent improvement) from different versions of the HWRF and GFS models for multiple storms in 2014-2016 and Hurricane Matthew in 2016, both with and without a gap in satellite coverage.

		All Observations		**Satellite Gap**	
Model		Multi Storm	Matthew 2016	Multi Storm	Matthew 2016
HWRF	V2015	–	–	5%	30%
	V2017	9%	–	–	–
	HEDAS	10%	–	–	–
GFS	V2015	–	8%	–	8%
	V2017	12%	28%	–	–

Forecast improvements were commonly largest at longer lead times. Peak benefits were often centered near a 3-day lead which, in many cases, corresponded to the greatest storm impacts

and/or the periods for which the observations were targeted. While gains at longer lead times might not directly support immediate actions such as watches and warnings, the observations demonstrate notable potential for improving challenging long-term forecasts which were often relatively poor for storm systems during the SHOUT campaigns.

The potential for positive forecast impact was most apparent for tropical cyclones, but benefits were also demonstrated for a winter storm that affected southern Alaska. Forecast improvements within GFS were observed for a wide range of forecast variables in the Alaska region impacted by the storm, but the magnitude of the gains was smaller than for the tropical storms and were not statistically significant given the sample size of a single storm. The similarity of improvements over different forecast variables suggests that the observed results are largely independent of the chosen forecast metric.

Not surprisingly, the impact of the GH observations was greatest on forecasts from model cycles during which the observations were collected and assimilated. Forecast benefits in those periods were especially large, particularly in the operational models. When results were averaged over multiple successive model cycles including periods in between flights or without observations, the impact was reduced but typically remained highly positive. The observations were observed to have a residual positive impact on subsequent model cycles, but the impact did decrease with time. A key benefit of the endurance of the GH is the ability to continually collect observations across more successive model cycles than other single aircraft. Less time is spent in transit than when sequential sampling is required from multiple aircraft.

Forecast improvements were also largest for the impacted and targeted regions. Analyses of other regions or times less severely affected by the sampled weather systems typically showed lesser forecast impacts and perhaps small degradations. Impacts on global forecast skill scores were largely neutral. A notable exception was the improvement in remote Pacific tropical cyclone tracks in the operational GFS model at NCEP/EMC resulting from assimilation of 2016 GH observations in the Atlantic. This suggested that the observations could potentially have more remote impact through procedures like satellite bias correction within the models. The lack of any broad global degradation from highly localized observations is positive on its own, as regional forecasts were improved without causing undue harm elsewhere.

As such, the potential value of GH-type observations may be specifically linked to the targeting of high impact weather events. This is consistent, however, with use in a satellite gap mitigation scenario where the aircraft would likely be flown preferentially in the event of severe weather as well as with current utilization of manned aircraft. A broader question concerns whether improving forecasts of particularly high impact events is important enough to justify possible small forecast degradations in other regions. This does not appear to be a major issue, however, at least in the operational models. It is also not an issue within tropical cyclone forecasting with HWRF which explicitly looks at the characteristics of a high-impact weather event. Initial application to tropical cyclones appears most justified.

Where explicitly evaluated, the impact of the GH observations was greater when added in the presence of a gap in environmental, polar-orbiting satellite observations. This supports the primary SHOUT project objective and the expectation that the GH could help supply coverage and upper level observations typically available only from satellites. The observation impact was almost uniformly positive when satellite observations were withheld. Within the regions of highest impact, forecast gains were less mixed, and positive forecast improvements were typically larger. Limited regional forecast degradations observed outside the highest impact areas when the GH data were added to the full observing system were largely eliminated when the observations were applied to a diminished observing system. While the operational NCEP/EMC results were not evaluated in the presence of a gap in satellite coverage, the ability to provide significant forecast improvements in the presence of the current full observing system implies that comparable or greater positive impacts should be expected in the absence of key components of the observing system.

The analyses were heavily weighted towards the impact of dropsonde observations. Studies examining the potential impact of the remote sensing payloads are still limited in scope. Initial experiments incorporating HAMSR, S-HIS, and HIRAD retrievals were all suggestive of positive impacts for forecasts of hurricanes and tropical storms, but more work is required before drawing meaningful conclusions. Work to this point was based upon retrievals as opposed to direct assimilation of radiances which could ultimately provide more benefit, consistent with satellite observations. The value of the endurance of the GH, particularly for mitigating any gaps in satellite coverage, would seemingly be enhanced through taking advantage of the continuity and coverage from remote sensors. Additional research involving the remote sensors is being proposed for the period following SHOUT.

Obtaining a large enough sample size of events to convincingly argue for significant, robust forecast gains is always challenging, particularly using a research asset like the GH. Through SHOUT and earlier research campaigns, the number of events sampled and analyzed has increased to the point where observed forecast improvements are being found to be statistically significant, at least for tropical cyclones, and meaningful results can begin to be inferred. The results within the operational modeling systems at NCEP/EMC are particularly positive in this regard. Moreover, the consistency of positive forecast benefits across models, targets, and metrics promotes increased confidence in the potential value of the observations.

Results from the OSSE studies are consistent with the OSEs in reflecting potential forecast benefits from addition of idealized GH-type observations. The primary guidance provided has been on sampling strategies with some insight into the relative merit of different observation variables. The hurricane OSSEs within HWRF-HEDAS specifically argued for enhanced sampling near storm centers. This finding was not, however, uniformly supported within other modeling systems where anecdotal evidence suggested enhanced value for some environmental observations. The impact of inner core observations may be enhanced within HEDAS where observations within 150 km of the storm center are not excluded as they are in the

operational HWRF system. This could argue for improved treatment of these observations in the future. The global OSSE studies supported the use of forecast sensitivity targeting guidance in flight path design in combination with sampling of key meteorological features. The value of temperature and wind observations were highlighted, but the studies yielded no conclusive guidance on optimum payload selection. Much can still be gained from further OSSEs, but realistic simulation of instrumental characteristics may continue to prove challenging.

Based in large part on SHOUT results, an action item was adopted at the 2017 Interdepartmental Hurricane Conference to support a further operational demonstration project evaluating GH dropsonde observations collected in August 2017 during the NASA-led Eastern Pacific Origins and Characteristics of Hurricanes (EPOCH) project. With the decision to remove restrictions on assimilation of GH dropsondes within GFS, the data were fully incorporated into the NOAA operational model suite, enabling a demonstration of its real-time impact. The action reflects growing support for the utility of GH observations and provides a next step towards demonstrating potential future operational utilization.

Overall, the results strongly support the hypothesis that GH-type observations can lead to consistent forecast improvements of high-impact weather events, particularly in the presence of a gap in key satellite observations. The capability to obtain tropical cyclone forecast improvements on the order of 10% or at least partially mitigate a gap in satellite coverage has been shown to exist today through deployment of dropsondes from the GH. Use of the GH could be a fundamental component of a gap-mitigation strategy in combination with other microsatellite missions also under evaluation. Questions regarding whether to utilize a GH-type UAS capability operationally in NOAA in the future will likely surround whether the forecast improvements are large or significant enough to justify the expense of operational usage of the platform. Results from the corresponding cost assessment (Kenul et al. 2018) must obviously be considered in partnership with these scientific findings. While the answers will ultimately be budgetary and programmatic as well as scientific, the potential for scientific benefit is broadly supported by the SHOUT analyses.

ACKNOWLEDGEMENTS

Many individuals and groups contributed to completing all of the components of the SHOUT project. SHOUT would not have been possible without the overall guidance and oversight from the Principal Investigator and Director of the NOAA UAS Program, Robbie Hood. All of the analyses presented in this report were conducted by teams from multiple laboratories including NCEP/EMC, AOML/HRD, and ESRL/GSD. We have attempted to cite the relevant contributors throughout the document, but the contributions from these groups cannot be overstated, particularly the leadership of Dr. Vijay Tallapragada, Dr. Altug Aksoy, and Dr. Lidia Cucurull. The studies would not have been possible without the extensive efforts that went into conducting the Global Hawk flights and collecting the data sets. Flight operations were supported by many groups including those at NASA AFRC, NOAA OMAO, NASA WFF, Northrop Grumman, the NOAA UAS Program Office, and the mission scientists from the SHOUT science team. The leadership efforts of Frank Cutler (NASA AFRC), CDR Jonathan Neuhaus (NOAA OMAO) and program managers Phil Kenul and John "JC" Coffey were especially significant. The critical data sets were produced by dedicated instrument teams led by Terry Hock from NCAR/EOL (AVAPS), Dr. Bjorn Lambrigtsen and Dr. Shannon Brown from NASA JPL (HAMSR), Dr. Gerald Heymsfield from NASA GSFC (HAMSR), Dr. Richard Blakeslee from NASA MSFC (LIP), Dr. Hank Revercomb from the University of Wisconsin (S-HIS), and Dr. Daniel Cecil from NASA MSFC (HIRAD). Input and guidance on SHOUT priorities was provided by an advisory group comprised of representatives across multiple NOAA line offices including Dr. Christopher Landsea, Carven Scott, Dr. Bill Ward, Dr. William Lapenta, Dr. Steven Goodman, Dr. Mitch Goldberg, Dr. Frank Marks, Dr. Robert Atlas, and Dr. John Cortinas. Valuable comments on the content of this report were provided by Robbie Hood, Dr. Lidia Cucurull, Dr. Altug Aksoy, and Dr. Peter Black. Formatting assistance was provided by Dawn Siemonsma from TriVector Services, Inc.

REFERENCES

Aksoy, A., S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F. Zhang, 2012: The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for high-resolution data: The impact of airborne Doppler radar observations in an OSSE. *Mon. Wea. Rev*., **140**, 1843-1862.

Aksoy, A., S. D. Aberson, T. Vukicevic, K. J. Sellwood, S. Lorsolo, and X. Zhang, 2013: Assimilation of high-resolution tropical cyclone observations with an ensemble Kalman filter using NOAA/AOML/HRD's HEDAS: Evaluation of the 2008-2011 vortex-scale analyses. *Mon. Wea.Rev*., **141**, 1842-1865.

Braun, S. A., and Coauthors, 2013: NASA's Genesis and Rapid Intensification Processes (GRIP) field experiment. *Bull. Amer. Meteor. Soc*., **94**, 345-363.

Braun, S. A., P. A. Newman, and G. M. Heymsfield, 2016: NASA's Hurricane and Severe Storm Sentinel (HS3) Investigation. *Bull. Amer. Meteor. Soc*., **97**, 2085-2102.

Christophersen, H., A. Aksoy, J. Dunion, and K. Sellwood, 2017: The impact of NASA Global Hawk unmanned aircraft dropwindsonde observations on tropical cyclone track, intensity, and structure: Case studies. *Mon. Wea. Rev*., **145**, 1817-1830.

Dole, R. M., et al., 2017: Advancing science and services during the 2015-16 El Nino: The NOAA El Niño rapid response field campaign. *Bull. Amer. Meteor. Soc*., in press.

Doyle, J. D., J. Moskaitis, P. G. Black, E. A. Hendricks, A. Reinecke, and C. M. Amerault, 2014: Multi- scale aspects of tropical cyclone predictability. American Geophysical Union Fall Meeting 2014, San Francisco, CA, December 2014.

Dunion, J. P., C. D. Thorncroft, and C. S. Velden, 2014: The tropical cyclone diurnal cycle of mature hurricanes. *Mon. Wea. Rev*., **142**, 3900-3919, DOI: 10.1175/MWR-D-13-00191.1.

Dunion, J. P., G. A. Wick, P. G. Black, J. Walker, 2018: Sensing Hazards with Operational Unmanned Technology: 2015-2016 Campaign Summary, Final Report. NOAA Tech Memo. OAR-UAS-001, 49 pp.

Hamill, T. M., F. Yang, C. Cardinalli, S. J. Majumdar, 2013: Impact of targeted winter storm reconnaissance dropwindsonde data on midlatitude numerical weather predictions. *Mon. Wea. Rev*., **141**, 2058-2065.

Kenul, P., J. Coffey, J. Walker, A. Roberts, and J. Huning, 2018: Sensing Hazards with Operational Unmanned Technology: Cost Study of Global Hawk Unmanned Aircraft System Operations for High Impact Weather Observations, Final Report. NOAA Tech Memo. OAR-UAS-002, 43 pp.

Kleist, D., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord, 2009: Introduction of the GSI into the NCEP Global Data Assimilation System. *Wea. Forecasting*, **24**, 1691-1705, doi:10.1175/2009WAF2222201.1.

Kren, A. C., L. Cucurull, and H. Wang, 2017: Impact of targeted UAS Global Hawk dropsonde

data on two high-impact weather events. *Mon. Wea. Rev*., submitted.

Lin, Y. 2011: GCIP/EOP Surface: Precipitation NCEP/EMC 4KM Gridded Data (GRIB) Stage IV Data. Version 1.0. UCAR/NCAR – Earth Observing Laboratory. http://data.eol.ucar.edu/dataset/21.093.

Nolan, D. S., R. Atlas, K. T. Bhatia, and L. R. Bucci, 2013: Development and validation of a hurricane nature run using the joint OSSE nature run and the WRF model. *J. Adv. Model. Earth Syst.*, **5**, 382- 405, doi:10.1002/jame.20031.

Peevey, T. R., J. M. English, L. Cucurull, H. Wang, and A. C. Kren, 2017: Improving winter storm forecasts with Observing System Simulation Experiments (OSSEs): Part 1, An idealized case study of three US storms. *Mon. Wea. Rev*., submitted.

Privé, N. C., Y. Xie, J. S. Wollen, S. E. Koch, R. Atlas, and R. E. Hood: 2013: Evaluation of the Earth Systems Research Laboratory's global Observing System Simulation Experiment system. *Tellus A*, **65**, 19011, http://dxdoi.org/10.3402/teuusa.v65i0.19011.

Sippel, J. A., S. A. Braun, F. Zhang, and Y. Weng, 2013: Ensemble Kalman filter assimilation of simulated HIWRAP Doppler velocity data in a hurricane. *Mon. Wea. Rev.*, **141**, 2683-2704.

Sippel, J. A., F. Zhang, Y. Weng, L. Tian, G. M. Hemsfield, and S. A. Braun, 2014: Ensemble Kalman filter assimilation of HIWRAP observations of Hurricane Karl (2010) from the unmanned Global Hawk Aircraft. *Mon. Wea. Rev.*, **142**, 4559-4580.

Tian, L., G. M. Heymsfield, A. C. Didlake, S. Guimond, and L. Li, 2015: Velocity-azimuth display analysis of Doppler velocity for HIWRAP. *J. Appl. Meteor. Climatol.*, **54**, 1792-1808.

Vömel, H., K. Young, and T Hock, 2016: Dropsonde Dry Bias, NCAR Technical Note, 4 pp.

Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-Based Ensemble-Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Mon. Wea. Rev.*, **141**, 4098-4117, doi:10.1175/MWR-D-12-00141.1.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev*., **130**, 1913–1924.

Wu, W.-S., D. F. Parrish, and R. J. Purser, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905-2916, doi:10.1175/1520- 0493(2002)130<2905:TDVAWS>2.0.CO;2.