On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream

Seethala Chellappan[1], Paquita Zuidema[2], James B Edson[3], Michael A Brunke[4], Gao Chen[5], Xiang-Yu Li[6], David Painemal[7], Claire E Robinson[8], Taylor Shingler[9], Michael Shook[10], Armin Sorooshian[11], Kenneth L Thornhill[5], Florian Tornow[12], Hailong Wang[6], Xubin Zeng[13], and Luke D. Ziemba[5]

[1]Rosenstiel School of Marine and Atmospheric Science
[2]University of Miami/RSMAS
[3]Woods Hole Oceanographic Institution
[4]University of Arizon, Tucson, AZ
[5]NASA Langley Research Center
[6]Pacific Northwest National Laboratory
[7]SSAI/NASA Langley Research Center
[8]Science Systems and Applications
[9]NASA Langley
[10]Langley Research Center
[11]University of Arizona
[12]Columbia University and NASA GISS
[13]The University of Arizona

November 21, 2022

Abstract

The warm Gulf Stream sea surface temperatures (SSTs) strongly impact the evolution of winter clouds behind atmospheric cold fronts. Such cloud evolution remains challenging to model. The Gulf Stream is too wide within the ERA5 and MERRA2 reanalyses, affecting the turbulent surface fluxes. Known problems within the ERA5 boundary layer (too-dry and too-cool with too strong westerlies), ascertained primarily from ACTIVATE 2020 campaign aircraft dropsondes and secondarily from older buoy measurements, reinforce surface flux biases. In contrast, MERRA2 winter surface winds and air-sea temperature/humidity differences are slightly too weak, producing surface fluxes that are too low. Reanalyses boundary layer heights in the strongly-forced winter cold-air-outbreak regime are realistic, whereas late-summer quiescent stable boundary layers are too shallow. Nevertheless, the reanalysis biases are small, and reanalyses adequately support their use for initializing higher-resolution cloud process modeling studies of cold-air outbreaks.

Posted on 21 Nov 2022 — CC-BY-NC 4 — https://doi.org/10.1002/essoar.10507559.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream

Seethala Chellappan[1], Paquita Zuidema[1], Jim Edson[2], Michael Brunke[3], Gao Chen[4], Xiang-Yu Li[5], David Painemal[4,6], Claire Robinson[4], Taylor Shingler[6], Michael Shook[6], Armin Sorooshian[3,7], Lee Thornhill[6], Florian Tornow[8,9], Hailong Wang[5], Xubin Zeng[3], Luke Ziemba[6]

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL
[2]Woods Hole Oceanographic Institution, Woods Hole, MA
[3]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ
[4]Science Systems and Applications, Inc., Hampton, VA
[5]Pacific Northwest National Laboratory, Richland, WA
[6]NASA Langley Research Center, Hampton, VA
[7]Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ
[8]Earth Institute, Columbia University, New York City, NY
[9]NASA Goddard Institute for Space Sciences, New York City, NY

Corresponding authors: Seethala Chellappan (seethala.chellappan@rsmas.miami.edu) and Paquita Zuidema (pzuidema@miami.edu)

Key Points:

- Reanalysis surface fluxes and boundary layers are representative of observations to first-order, sufficient for higher-resolution model initialization.
- Reanalyses represent the Gulf Stream more broadly than is seen in nature, contributing to turbulent flux and boundary layer biases.
- Previously-noted thermodynamic and dynamic biases reinforce (ERA5) or compensate (MERRA2) surface fluxes but support realistic winter boundary layer heights.

Abstract

The warm Gulf Stream sea surface temperatures (SSTs) strongly impact the evolution of winter clouds behind atmospheric cold fronts. Such cloud evolution remains challenging to model. The Gulf Stream is too wide within the ERA5 and MERRA2 reanalyses, affecting the turbulent surface fluxes. Known problems within the ERA5 boundary layer (too-dry and too-cool with too strong westerlies), ascertained primarily from ACTIVATE 2020 campaign aircraft dropsondes and secondarily from older buoy measurements, reinforce surface flux biases. In contrast, MERRA2 winter surface winds and air-sea temperature/humidity differences are slightly too weak, producing surface fluxes that are too low. Reanalyses boundary layer heights in the strongly-forced winter cold-air-outbreak regime are realistic, whereas late-summer quiescent stable boundary layers are too shallow. Nevertheless, the reanalysis biases are small, and reanalyses adequately support their use for initializing higher-resolution cloud process modeling studies of cold-air outbreaks.

Plain Language Summary

The Gulf Stream is a narrow band of warm water to the east of continental north America. As air moves eastward off of the continent, the warm ocean temperatures transfer moisture and heat that help develop and modify the marine low clouds. This transfer, particularly during cold-air outbreaks present significant modeling challenges that contribute uncertainty to temperature projections for a world with more carbon dioxide. Simulations seeking to represent the details of such shallow clouds must rely on initializations and forcings that originate from coarser-resolution renalyses. Here we explore how well two major reanalyses and a commonly-used flux product represent these fluxes and the boundary layer, using ocean buoy measurements and new *in-situ* observations from an aircraft campaign. We find that the reanalyses are adequate for the purpose of initializing higher-resolution modeling of the cold-air outbreak clouds. In particular, the winter boundary layer heights are realistic. These heights are important for capturing winter cloud-environmental interactions correctly. Late summer boundary layers are too shallow. Known biases do remain present, and impact the surface flux errors differently in the two reanalyses examined.

1. Introduction

A prominent feature of the northwest Atlantic is the Gulf Stream, a western boundary current transporting warm waters to the north. The Gulf Stream is up to 10°C warmer than the surrounding waters (Fig. 1), fueling atmospheric convection. During the off-summer months, westward-moving mid-latitude synoptic disturbances exchange warm, low-latitude air with colder air. In this situation, the air-sea interaction becomes exceptional. The evolution of the low marine clouds in the post-frontal regions of strong subsidence is described in Grossman and Betts (1990); Kolstad et al., (2009); Liu et al., (2014); Fletcher et al., (2016a, 2016b); McCoy et al., (2017), and Painemal et al., (2021). These clouds remain challenging to model (Skyllingstad and Edson, 2009; Field et al., 2014; Abel et al., 2017). Leading questions remain realistic representations of the roll cloud circulations (Honnert et al, 2020), and the correct partitioning between the liquid and ice phases (Mulmenstadt et al., 2014; Field and Heymsfield, 2014). The latter contributes to the cloud feedback uncertainty in climate models (Zelinka et al., 2020; Sherwood et al., 2020).

The sharp increase in the sea surface temperature (SST) of the Gulf Stream is particularly noticeable in winter. Turbulent fluxes can exceed 1000 W m^{-2} during cold-air outbreak events (Bane and Osgood, 1989; Marshall et al., 2009; Biggore et al., 2013). The adjustment of the boundary-layer air temperature and humidity to the underlying surface can establish a thermally-direct circulation, in which horizontal pressure and boundary-layer height gradients drive a surface wind convergence on the warmer flank of the Gulf Stream (Minobe et al., 2008; Liu et al., 2014; Plagge et al., 2016). The increase in surface winds is aided by a downward transfer of momentum and the shear mediates an adjustment to the altered boundary layer stratification as well (Small et al., 2008).

The northwest Atlantic is a strategic environment for improving the understanding and modeling of cold-air outbreaks (CAOs) through observations, aided by the proximity to the eastern north American seaboard. The characterization and process modeling of CAOs is a key objective of the NASA Earth Venture Suborbital-3 Aerosol-Cloud-meTeorology Interactions oVer the western Atlantic Experiment (ACTIVATE; Sorooshian et al., 2019). The process modeling activities rely on reanalysis data for initialization (Tornow et al., 2021; Li et al., 2021), and include subsequent nudging to above-inversion values (Tornow et al., 2021). The ACTIVATE winter 2020 campaign included two dropsonde circles to explicitly derive vertical

velocities following Bony and Stevens, (2019), out of concern that reanalysis-derived vertical motion might not be adequate for model forcings. Reanalysis fluxes provide a tempting alternative to derived *in-situ* fluxes requiring long, low-altitude level legs that compete with the gathering of new cloud microphysical information. Reanalyses, in combination with satellite datasets, also provide useful longer-term context (Painemal et al., 2021).

This study addresses the following two questions: 1) How accurate are the surface fluxes from the latest major reanalyses (the fifth-generation ECMWF (ERA5) and Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA2)) and the Objectively Analyzed air-sea Heat Fluxes (OAFLUX) in the presence of wintertime CAOs over the Gulf Stream? This extends prior assessments based on coarser-resolution products (Moore and Renfrew, 2002; Jin and Yu, 2013. 2) Can ERA5 and MERRA2 provide a realistic depiction of the Gulf Stream-affected boundary layer? We rely on buoy and ACTIVATE dropsonde data for reference. The questions are relevant beyond the scope of the ACTIVATE campaign, and recognize the challenges inherent to representing strong air-sea coupling events. Reanalysis products can be an alternative to observations in weather and climate studies and are also applied to climate model assessments – sometimes without fully understanding their performance.

2. Datasets and Method

Buoy measurements including direct covariance buoyancy fluxes from the CLIMODE (Climate Variability and Predictability Mode Water Dynamics Experiment; Marshall et al., 2009) campaign provide absolute reference values. The CLIMODE buoy was situated at 38°N, 65°W (Fig. 1), with the data from the 12 months of year 2006 incorporated into this study. The buoy was to the north of the Gulf Stream in Febuary-March, 2006, and in August-September, 2006, within and south of the Gulf Stream (Fig. S1). Twenty-minute averages of temperature, relative humidity and wind speed were measured at about 3m above the waterline and calibrated, with drifts and biases corrected (Weller et al., 2012; Bigorre et al., 2013), and equated to the 2 and 10 meter values available within reanalyses (T_{2m}, RH_{2m}, WS_{10m}). Both a surface-skin SST (*SST_skin*), to which the surface fluxes are responsive, and a sub-surface foundation SST (*SST_foundation*) are measured.

The ACTIVATE campaign sampled on both sides of the northern Gulf Stream SST gradient (Fig. 1). During February-March, the Gulf Stream meandered to the north, more

noticeable west of 73°W, altering the local SST by more than 4K in places (Fig. 1). The more quiescent synoptic conditions in August-September over a more uniform area support an assessment of boundary-layer depictions less influenced by strong air-sea interactions. Dropsondes were launched from the UC-12 King Air flying at approximately 9 km. Of the 13 King Air flight days in February-March, 2020, 8 coincided with visually-identified CAOs, encompassing 43 of the 59 winter dropsondes (Table S1 lists the individual flight days and their designation as CAO/non-CAO days). Dropsonde circles provide intensive sampling of two CAOs and are the subject of detailed simulations (Li et al., 2021). The 18 August-September flight days include 3 (weaker) CAOs and deployed 107 dropsondes total**.** Neither the buoy nor the dropsonde data are assimilated in either reanalysis or flux product.

The satellite-derived Group for High-Resolution SST (GHRSST), at 9 km spatial resolution, provides spatial context and input to bulk flux calculations. The provided *SST_foundation* at 1m depth is derived from the measured *SST_skin* using a diurnal model.

The ERA5 Reanalysis is arguably the publicly-available reanalysis with the most sophisticated depiction of the cloudy boundary layer. ERA5 has a horizontal grid spacing of 31 km, 137 vertical levels, and an hourly temporal resolution. A systematic bias in the partitioning of ERA5's global wind kinetic energy, with an excessive mean zonal flow coupled with weak meridional flow, is attributed to difficulty in representing high-frequency transient atmospheric events (Rivas and Stoffelen, 2019). The mid-latitude boundary layer maintains a cold and dry bias (Hersbach et al., 2020). The assimilated SST product prior to 2007 was based on the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST2; Titchner & Rayner, 2014), a 0.25°x0.25° pentad product too coarse to resolve Gulf Stream SST gradients well (Chelton and Risien, 2016). After 2007 it was based on the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012). OSTIA is produced daily on a higher-resolution 0.05^{o}x0.05^{o} grid, and includes satellite microwave measurements, which are less sensitive to the presence of cloud. Systematic differences between the ERA5 SST still remain from other SST climatologies, attributed to spatial resolution (Hersbach et al., 2020). ERA5 provides both *SST_skin* and *SST_foundation*. The 2020 ERA5 *SST_skin* is typically cooler than the *SST_foundation* by 0.2-0.4K (Fig. S2) because of infrared cooling (Fairall et al., 1996; Minnett et al., 2019), with little diurnal warming apparent during the windier, overcast time periods primarily represented here.

MERRA2 spatial resolution is 0.625° x 0.5° longitude by latitude, with 72 vertical levels, and a three-hour temporal resolution, although the surface latent and sensible heat fluxes and lowest-model-level meteorological variables are available at a higher one-hour interval. Previous comparisons to soundings over the Beaufort Sea indicate a MERRA2 warm bias near the surface, and no humidity bias (Rozenhaimer et al., 2018). MERRA2 winds appear too weak, by up to 2 m s^{-1}, in the ACTIVATE region within Molod et al., (2015). MERRA2 only provides *SST_skin*. OAFLUX, at a 1°x1° spatial resolution, provides further insight into the impact of spatial resolution.

Surface fluxes are calculated from the individual dropsondes using the COARE v3.5 parameterization. The COARE v3.0 bulk parameterization values, applied to the buoy T_{2m}, q_{2m} and $WS_{10m,}$ compare well against the CLIMODE buoy direct covariance buoyancy fluxes (Edson et al., 2013; Bigorre et al., 2013), with modifications at wind speeds > 12 m s^{-1} based on the CLIMODE measurements leading to the COARE v3.5 bulk flux algorithm (Edson et al., 2013). Overall this indicates that if the near-surface parameters are known, then the buoyancy fluxes can be estimated with little bias, even within CAOs. Dropsonde inter- or extrapolation provides T_{2m} and q_{2m} and an *SST_skin* is estimated as GHRSST *SST_foundation* + (ERA5 *SST_skin* – ERA5 *SST_foundation*).

The reanalysis comparisons to the dropsondes are nearest in time and space. Instantaneous values captured by the dropsondes from up- and downdrafts will increase the variability of the dropsonde-calculated fluxes beyond those of the coarser-resolution reanalysis fluxes. The comparisons to the CLIMODE buoy data are of the daily- and monthly-mean values, ignoring diurnal variations in wind speed (Dai and Deser, 1997) and near-surface humidity (Clayson and Edson, 2019). Further details can be found in the Supplement.

4. Surface flux representations

a. 2006 CLIMODE

The buoy was located within an SST gradient of approximately 8K over a mere 100 km in January-April of 2006 (Fig. S1). This constitutes a challenging regime for any reanalysis. The monthly-mean SSTs exceed buoy values by up to 5K during February-April, indicating reanalyses depictions of the Gulf Stream that are broader than in nature (Fig. 2 and Fig. S3),

more noticeable as spatial-resolution degrades. Consistent with this, the reanalyses T_{2m} are too warm, and the saturated specific humidity (q_s) too high. The winds are too strong, which physically can be related to the too-warm ocean surface (Small et al., 2008). Monthly-mean reanalyses and OAFLUX buoyancy fluxes differ significantly from the CLIMODE direct covariance values (Fig. S3), most notably in February-March, when buoyancy fluxes exceed the buoy values by 40 to 60 W m^{-2} (see also Table S2), an overestimate of >80%. Since the bulk flux calculations are validated (Edson et al., 2013), the root of the reanalyses flux biases must be their SST representation. This conclusion is in line with Jin and Yu (2013), extended here to newer reanalyses possessing a higher spatial resolution.

Daily-mean differences between the reanalyses/product and CLIMODE buoy values (Fig. 2, Fig. S4 and Table S2) clarify the atmospheric consequences of the SST misrepresentations. The overestimated SST skews reanalysis wind speeds to positive values (Fig. 2f). The ERA5 *SST_foundation* and T_{2m} deviate the least from the buoy values (Fig. 2a, b), even though the MERRA2 *SST_skin* values should in theory be cooler (Fig. S2). Both reanalyses match the buoy wind speeds well (Fig. 2f, Table S2). The surface q_s is elevated for all reanalyses/product, as expected. A dry bias in ERA5's q_{2m} contrasts with a moist bias for MERRA2's q_{2m}, both by about 1 g kg^{-1}. In combination, the air-sea thermodynamic differences are smaller for MERRA2 on most days, compared to ERA5 (Fig. 2g and h, Table S2), compensating for MERRA2's poorer T_{2m} (Fig. 2a). This allows the MERRA2 buoyancy fluxes to ultimately compare better to the CLIMODE values, than the ERA5 values (Fig. 2e). Overall, this comparison suggests ERA5 provides a more accurate depiction of the Gulf Stream near-surface meteorology, likely in part because of an improved resolution, but compensations within MERRA2 model physics may be improving the fluxes, if for the wrong reasons. OAFLUX, with the coarsest resolution, has flux, SST and wind speed values that diverge the most of the three products from CLIMODE buoy values (Fig. S3; Fig. 2b, e, f; Table S2).

During the summer months, the SST differences are smaller and more evenly distributed about zero (Fig. 2j and S4). Buoyancy fluxes remain consistently overestimated (Fig. 2m), most noticeable by ERA5 because of its too-cool T_{2m} and too-dry (by 1-2 g/kg) q_{2m}. The too-dry ERA5 q_{2m} bias for both seasons indicates a common bias source. In contrast, a too-warm T_{2m} in winter and too-cool T_{2m} in summer suggests differing underlying causes.

b. 2020 ACTIVATE

ERA5 *SST_foundation* exceed GHRSST values by up to 2K during February-March of 2020 at the Gulf Stream boundaries, and a slight underestimation of the cooler southward-flowing coastal Labrador Current temperatures is also evident (Fig. 1d). The ERA5 SST bias exists even though the assimilated SST product is of a similar (slightly finer) spatial resolution as GHRSST (0.05° versus 9 km), reflecting the coarsening needed to match the ERA5 resolution of 31 km (Hersbach et al., 2020).

The hourly-mean ERA5 reanalysis sensible and latent heat fluxes overestimate during the more severe CAOs, by up to 100 W m^{-2} (Fig. 3a) or more for the latent heat fluxes (Fig. 3b). The cause is most clearly linked to wind speed overestimates. In contrast, the MERRA2 fluxes always underestimate, because of underestimates in the air-sea temperature (Fig. 3c) and humidity (Fig. 3d) differences, and weaker MERRA2 wind speeds (Fig. 3e). Differences in temporal/spatial resolution (instantaneous versus hourly-mean values over a larger spatial domain) seem unlikely to explain the MERRA2 underestimates, given that these are systematic biases, and instead point to a near-surface boundary layer that is too close in thermodynamic equilibrium with the ocean.

In August-September 2020, both reanalyses slightly underestimate the fluxes relative to those calculated from the dropsondes, with ERA5 performing better than MERRA2 (Table S3). Air-sea humidity differences are more realistically captured by ERA5 than by MERRA2. MERRA2, similar to the winter months, consistently underestimates all inputs into the flux calculations, although the biases are small (Table S3).

5. Thermodynamic Vertical Structure

Figure 4 compares the ERA5 and MERRA2 vertical atmospheric structure to the dropsonde-derived mean potential temperature (θ), *RH* and *q*, and wind speed (total, zonal and meridional) for February-March and August-September of 2020, while Fig. S5 indicates the differences more explicitly. The reanalyses capture the main features of the lower tropospheric structure. Consistent with the near-surface analysis, the wintertime ERA5 boundary layer is slightly too cold and too dry, with *RH* and *q* averages indicating underestimates of 5% (ranging up to 30%) and 0.5 g kg^{-1} (ranging up to 2 g kg^{-1}), respectively. Locations with ERA5-SST – GHRSST > 1K reveal mean ERA5 boundary layer θ profiles that are 0.2K warmer, ranging up to

1 to 2K. Where ERA5-SST – GHRSST < -1K, the ERA5 θ profiles are almost 1K cooler than the dropsonde values (inset plot in Fig. 4a).

The mean ERA5 wind biases are small, with a slight overestimation (1 m s^{-1}, or 10%) that primarily comes from the zonal component. The lower free troposphere in ERA5 does not fully resolve the observed structure (e.g., the elevated moisture layer between 800-750 hPa), but the main inversion top at approximately 850 hPa, identified using an *RH* threshold, is adequately captured. The winter MERRA2 thermodynamic structure compares more closely to the *in-situ* values (seen more clearly in Fig. S5). Interestingly, the sign of the MERRA2 wind bias contrasts with that from ERA5. An underestimate of the near-surface zonal winds increases with altitude, suggesting a downward momentum transport may explain the near-surface bias.

The ERA5 wind biases are smaller in August-September, while ERA5 θ remains depressed by 0.2 - 0.3 K (ranging up to 3K) and q also remains biased low, by up to 0.5 g kg^{-1}. These compensate to generate realistic *RH* values near the surface (Fig. S5). Specific humidity underestimates above the surface-based mixed layer are larger for MERRA2 than ERA5, permeating into the relative humidity. Both ERA5 and MERRA2 struggle with capturing the cloud layer between 900-800 hPa (Fig. 4h inset). This is not linked to a pronounced bias in the winds for ERA5, while MERRA2 winds are clearly too weak above 1 km.

In contrast to ERA5, the *RH* and q-mean MERRA2 profiles agree well with the dropsondes during February-March, while the zonal winds are consistently weaker, by 1-2 m s^{-1} near the surface, increasing (mostly) with altitude. During the late summer, MERRA2 is more likely to be drier within 0.4-2.0 km than the *in-situ* measurements, indicating the critical relative humidity threshold for cloud production may be set too low then (Molod et al., 2015).

During February-March, both ERA5 and MERRA2 capture the inversion height of approximately 1.7 km reasonably well (estimated from the *RH* profiles). During August-September, the inversion is naturally lower, at approximately 1.1 km (similarly estimated). Both reanalyses often fail to capture the cloud layer in late summer. The mean MERRA2 boundary layer height is lower than that from ERA5, with a drier cloud layer. The lower boundary layer height is even more pronounced after the few September CAO cases are excluded (not shown), indicating the issue may be a similar difficulty in representing stable boundary layers as for ERA5. The MERRA2 winds, both zonal and meridional, are also weaker.

5 Discussions and Conclusions

Hersbach et al. (2020) note that the too-dry ERA5 boundary layer, evident in all seasons and in both the CLIMODE and ACTIVATE comparisons, coincides with a warming of the lower troposphere and with the advent of microwave imagers. These are shown to warm and dry ERA5 at 850 hPa over the ocean (Geer et al., 2017). The exact mechanisms do not yet appear to be known (Hersbach et al., 2020); the assimilation of microwave radiances, although providing additional information, can nevertheless not fully constrain the thermodynamic profile (e.g., Pincus et al., 2017; Zhang et al., 2018), introducing understandable trade-offs.

The too-strong wintertime westerlies generate an anomalous wind convergence near the surface (Rivas and Stoffelen, 2019), which should act to raise the boundary layer height, all else being equal. The reasonable depiction of the ERA5 wintertime boundary layer depth, whereas the late summer cloudy boundary layer is too shallow, may reflect a conscious choice at numerical weather prediction centers to artificially enhance the turbulent diffusion in stable conditions, towards improving the depiction of synoptic cyclones (Sandu et al., 2013). The ACTIVATE campaign is selectively sampling CAO conditions during its winter campaigns, for which the ERA5 turbulent diffusion choices are optimized. The late summer time period, when the north Atlantic sea level pressure high extends further west (Painemal et al., 2021), provides conditions in which both reanalyses have more difficulty in maintaining a cloudy stable layer. The artificial enhancement in the ERA5 diffusion parameters was also intended to improve a near-surface cold temperature bias (by encouraging the entrainment of warmer air aloft); we find a small (~0.2K) ERA5 cold temperature bias still remains during both winter and late summer.

The length scale of ocean mesoscale eddies is 20-30 km at the latitude of the Gulf Stream, set to first-order by the Rossby radius. ERA5 possesses the horizontal grid spacing best able to represent the majority of the ocean mesoscale activity of the three products examined, though still missing the smallest eddies. A wintertime western boundary current that is too wide in the reanalyses could imply that the boundary layer adjustment for air coming from the west might be affected earlier within the reanalysis than in nature. CAO air flows first over cooler coastal waters north of 35°N generated by the Labrador Current, whose ERA5 reanalysis temperatures are too cool, potentially further energizing the adjustment process of the boundary layer to the warmer Gulf Stream waters. In addition, all of the biases in ERA5 contribute to

exaggerating the surface heat fluxes during CAOs, which will also contribute to elevating the inversion. Perhaps because of these characteristics, the wintertime ERA5 boundary layer depth, thermodynamic and dynamic structure is broadly representative of the observations. ERA5 vertical motion fields have also been shown to compare well to those derived from the two dropsonde circles (Li et al., 2021), lending further confidence in the ability of ERA5 to depict the strongly-forced cold-air outbreak regime. A correct boundary layer depth is a critical parameter for shallow clouds, as the depth affects the coupling to the ocean surface. The robustness of the bias in ERA5 q suggests an observationally-determined correction factor could be applied, improved as more dropsonde data become available. A similar approach could be adopted for the too-cool ERA5 temperature bias. For those LES studies assuming a Lagrangian perspective, additional ACTIVATE dropsonde data will also support analysis of how much reanalysis profiles deviate from observations as a function of distance from shore. Interestingly, the biases in the MERRA2 reanalysis are different from those in ERA5, tending to too-weak surface fluxes. Nevertheless the boundary layer depth depiction is similar: approximately realistic during the winter, and too shallow during the summer. The momentum transport to the surface by ERA5 can increase the surface wind speed, deepening the boundary layer more quickly and encouraging a faster cloud transition, than in nature (Saggiaroto et al., 2020). Future work will incorporate space-based lidar and radar data and *in-situ* measurements to evaluate the CAO cloud structure evolution over the Gulf Stream.

Acknowledgments and Data Availability

This work was supported by NASA grant 80NSSC19K0390 in support of ACTIVATE, a NASA Earth Venture Suborbital-3 (EVS-3) investigation funded by NASA's Earth Science Division and managed through the Earth System Science Pathfinder Program Office. The Pacific Northwest National Laboratory (PNNL) is operated for the US Department of Energy (DOE) by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830. Microwave OI SST data are produced by Remote Sensing Systems, available at www.remss.com. ERA5 data are available through https://cds.climate.copernicus.eu/. MERRA-2 data are available at https://disc.gsfc.nasa.gov/. CLIMODE buoy measurements are available at http://www.opal.sr.unh.edu/data/airsea_flux.shtml. The OAFlux data are available through http://oaflux.whoi.edu/.

References

Abel, S. J., Boutle, I. A., Waite, K., Fox, S., Brown, P. R. A., Cotton, R., Lloyd, G., Choularton, T. W., & Bower, K. N. (2017). The Role of Precipitation in Controlling the Transition from

Stratocumulus to Cumulus Clouds in a Northern Hemisphere Cold-Air Outbreak, *Journal of the Atmospheric Sciences*, *74*(7), 2293-2314.

Bane, J. M., Jr., & Osgood, K. E. (1989). Wintertime air-sea interaction processes across the Gulf Stream. *Journal of Geophysical Research*, 94 (C8), 10 755–10 722. doi: 10.1029/JC094iC08p10755.

Bigorre, S., Weller, R. A., Lord, J., Edson, J. B., & Ware, J. D. (2013). A surface mooring for air–sea interaction research in the Gulf Stream. Part 2: Analysis of the observations and their accuracies. *Journal of Atmosheric and Oceanic Technology*, 30, 450–469, doi:10.1175/ JTECH-D-12-00078.1.

Bony, S., & Stevens, B. (2019). Measuring Area-Averaged Vertical Motions with Dropsondes, *Journal of the Atmospheric Sciences*, *76*(3), 767-783

Bosilovich, M. G. and Coauthors. (2015). *MERRA-2: Initial evaluation of the climate. Technical Report Series on Global Modeling and Data Assimilation*, Vol. 43, NASA Tech. Rep. NASA/TM– 2015–104606, 139 pp. [Available online at https://gmao.gsfc. nasa.gov/pubs/docs/Bosilovich803.pdf.]

Chelton, D. B., & Risien, C.M. (2016). *Zonal and Meridional Discontinuities and Other Issues with the HadISST1.1 Dataset*. Oregon State University, Corvallis, Oregon, Technical Report. doi: 10.13140/RG.2.1.4503.0168.

Clayson, C. A., & Edson, J. B. (2019). Diurnal surface flux variability over western boundary currents. *Geophysical Research Letters*, 46, 9174–9182. Doi: 10.1029/2019GL082826

Dai, A., & Deser, C. (1997). Diurnal and semidiurnal variations in global surface wind and divergence fields. *Journal of Geophysical Research*, 104, 31,109–31,125

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, 116, 140–158, doi:10.1016/j.rse.2010.10.017.

Edson, J. B., Jampana, V., Weller, R.A., Bigorre, S.P., Plueddemann, A.J., Fairall, C.W., Miller, S.D., Mahrt, L., Vickers, D., & Hersbach, H. (2013). On the exchange of momentum over the open ocean. *Journal of Physical Oceanography*, 43, 1589–1610. doi:10.1175/ JPO-D-12-0173.1.

Fairall, C.W., Bradley, E.F., Godfrey, J.S. , Edson, J.B., Young, G.S. and Wick, G.A. (1996) Cool skin and warm layer effects on the sea surface temperature. *J. Geophys. Res., 101,* 1295-1308.

Fairall C. W., Bradley, E., Hare, J., Grachev, A., & Edson, J. (2003). Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm. *Journal of Climatology*, 16, 571–591. doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2

Field, P. R., Cotton, R. J., McBeath, K., Lock, A. P., Webster, S., & Allan, R. P.(2014). Improving a convection-permitting model simulation of a cold air outbreak. *Quarterly Journal of Royal Meteorological Society*, 140, 124–138, https://doi.org/10.1002/qj.2116.

Field, P. R., & Heymsfield, A. J. (2015). Importance of snow to global precipitation. *Geophysical Research Letter*, 42, 9512–9520, doi:10.1002/2015GL065497.

Fletcher, J. K., Mason, S. L., & Jakob, C. (2016a). The climatology, meteorology, and boundary layer structure of marine cold air outbreaks in both hemispheres. *Journal of Climate*, 29, 1999–2014, https://doi.org/10.1175/JCLI-D-15-0268.1.

Fletcher, J. K., Mason, S. L., & Jakob, C. (2016b). A climatology of clouds marine cold air outbreaks in both hemispheres. *Journal of Climate*, 29, 6677–6692, https://doi.org/10.1175/JCLI-D-15-0783.1.

Geer, A.J., Baordo, F., Bormann, N., English, S.J., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. & Lupu, C. (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, 143, 3189–3206. doi:10.1002/qj.3172

Grossman, R. L.& Betts, A. K. (1990). Air–sea interaction during an extreme cold air outbreak from the eastern coast of the United States. *Monthly Weather Review*, 118, 324–342.

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global analysis. *Quarterly Journal of Royal Meteorological Society*, 146, 1999-2049. doi:10.1002/qj.3803

Honnert, R., Efstathiou, G., Beare, R., Ito, J., Lock, A., Neggers, R., et al. (2020). The atmospheric boundary layer and the "gray zone" of turbulence: A critical review.*Journal of Geophysical Research: Atmospheres*, 125, e2019JD030317. doi.org/10.1029/2019JD030317

Jin, X., and L. Yu (2013), Assessing high-resolution analysis of surface heat fluxes in the Gulf Stream region, *J. Geophys. Res. Oceans*, 118, 5353–5375, doi:10.1002/jgrc.20386

Kolstad, E.W., Bracegirdle, T. J., & Seierstad, I. A. (2009). Marine cold-air outbreaks in the North Atlantic: Temporal distribution and associations with large-scale atmospheric circulation. *Climate Dynamics*, 33, 187–197, https://doi.org/10.1007/s00382-008-0431-5.

Li, X.-Y., Wang, H., Chen, J., Endo, S., George, G., Cairns, B., et al., (2021). Large-eddy simulations of marine boundary-layer clouds associated with cold air outbreak during the ACTIVATE campaign – Part 1: Case setup and sensitivities to large-scale forcings. *Journal of Atmospheric Science*. Submitted.

Liu, J., Xie, S., Norris, J. R., & Zhang, S. (2014). Low-Level Cloud Response to the Gulf Stream Front in Winter Using CALIPSO. *Journal of Climate*, *27*(12), 4421-4432.

Marshall, J., Andersson, A., Bates, N., Dewar,W., Doney, S., Edson, J., et al., (2009). The CLIMODE field campaign: Observing the cycle of convection and restratification over the Gulf Stream. *Bull. Amer. Meteor. Soc., 90*, 1337-1350.

Masunaga, R., Nakamura, H., Miyasaka, T., Nishii, K., & Tanimoto, Y. (2015). Separation of Climatological Imprints of the Kuroshio Extension and Oyashio Fronts on the Wintertime Atmospheric Boundary Layer: Their Sensitivity to SST Resolution Prescribed for Atmospheric Reanalysis. *Journal of Climate*, 28, 1764–1787. doi:10.1175/JCLI-D-14-00314.1.

McCoy, I., Wood, R., & Fletcher, J. K. (2017). Identifying meteorological controls on open and closed mesoscale cellular associated with marine cold air outbreaks. *Journal of Geophysical Research,* 122, 11,678–11,702, https://doi.org/10.1002/2017JD027031, 2017

Minnett, P. J., A. Alvera-Azcárate, T.M. Chin, G.K. Corlett, C.L. Gentemann, I. Karagali, et al., (2019). Half a century of satellite remote sensing of sea-surface temperature, *Remote Sensing of Environment*, 233, doi: 10.1016/j.rse.2019.111366.

Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S.P., & Small, R.J. (2008). Influence of the Gulf Stream on the troposphere. *Nature*, 452, 206–209, doi: 10.1038/nature06690.

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development,* 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Moore, G. W. K., & Renfrew, I.A. (2002). An assessment of the surface turbulent heat fluxes from the NCEP–NCAR reanalysis over the western boundary currents. *Journal of Climate*, 15, 2020–2037. doi: 10.1175/1520-0442(2002)015<2020:AAOTST>2.0.CO;2

Mülmenstadt, J., Sourdeval, O., Delano., J., & Quaas. J. (2015). Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals. *Geophysical Research Lett*er 42, 6502–6509, doi:10.1002/2015GL064604.

Pincus, R., Beljaars, A., Buehler, S.A., Kirchengast, G., Ladstaedter, F., Whitaker, J. S. (2017). The Representation of Tropospheric Water Vapor Over Low-Latitude Oceans in (Re-)analysis: Errors, Impacts, and the Ability to Exploit Current and Prospective Observations. *Surv Geophys* 38**,** 1399–1423. doi:10.1007/s10712-017-9437-z

Plagge, A., Edson, J. B., & Vandemark, D. (2016). In Situ and Satellite Evaluation of Air–Sea Flux Variation near Ocean Temperature Gradients. *Journal of Climate*, *29*(4), 1583-1602.

Painemal, D., Corral, A. F., Sorooshian, A., Brunke, M. A., Seethala, C.,, Gorooh, V. A. et a., (2021). An Overview of Atmospheric Features Over the Western North Atlantic Ocean and North American East Coast – Part 2: Circulation, Boundary Layer, and Clouds. *Journal of Geophysical Research,* 126, e2020JD033423. doi:10.1029/2020JD033423

Remote Sensing Systems (REMSS) (2008). *GHRSST Level 4 MW_IR_OI Global Foundation Sea Surface Temperature analysis version 5.0 from REMSS (GDS versions 1 and 2)*. NOAA

National Centers for Environmental Information. doi: gov.noaa.nodc:GHRSST-MW_IR_OI-REMSS-L4-GLOB

Reynolds, R.W. & Smith, T.M. (1994). Improved Global Sea Surface Temperature Analysis using Optimum Interpolation. *Journal of Climate*, 7, 929-948. doi: 10.1175/1520-0442(1994)007

Rivas, M. B., & Stoffelen, A. (2019). Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Science*, 15, 831-851. doi: 10.5194/os-15-831-2019

Rozenhaimer, M., Barton, N., Redemann, J., Schmidt, S., LeBlanc, S., Anderson, B., et al. (2018). Bias and sensitivity of boundary layer clouds and surface radiative fluxes in MERRA-2 and airborne observations over the Beaufort Sea during the ARISE campaign. *Journal of Geophysical Research,* 123, 6565–6580. Doi:10.1029/2018JD028349

Saggiorato, B., Nuijens, L., Siebesma, A. P., de Roode, S., Sandu, I.,& Papritz, L. (2020). The influence of convective momentum transport and vertical wind shear on the evolution of a cold air outbreak. *Journal of Advances in Modeling Earth System*, 12, https://doi.org/10.1029/2019MS001991.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58, e2019RG000678. https://doi.org/10.1029/2019RG000678.

Skyllingstad, E. D., & Edson, J. B. (2009). Large-eddy simulation of moist convection during a cold-air outbreak over the Gulf Stream. *Journal of the Atmospheric Sciences*, 66(5), 1274–1293. https://doi.org/10.1175/2008JAS2755

Small, R. J., & Coauthors. (2008). Air–sea interaction over ocean fronts and eddies. *Dynamics of Atmosphere and Oceans*, 45, 274–319, doi:10.1016/j.dynatmoce.2008.01.001.

Sorooshian, A., Anderson, B., Bauer, S. E., Braun, R. A., Cairns, B., Crosbie, E., et al., (2019). Aerosol–cloud–meteorology interaction airborne field investigations: Using lessons learned from the U.S. West Coast in the design of ACTIVATE off the U.S. East Coast. *Bulletin of the American Meteorological Society*, 100(8), 1511– 1528. https://doi.org/10.1175/BAMS-D-18-0100.1.

Titchner, H. A., & Rayner, N.A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research Atmospheres, 119,* 2864-2889. doi: 10.1002/2013JD020316.

Tornow, F., Ackerman, A. S., & Fridlind, A. M. (2021). Preconditioning of overcast-to-broken cloud transitions by riming in marine cold air outbreaks. *Atmospheric Chemistry and Physics Discussions*, https://doi.org/10.5194/acp-2021-82.

Weller, R.A., Bigorre, S.P., Lord, J., Ware, J.D., & Edson, J.B. (2012). A Surface Mooring for Air–Sea Interaction Research in the Gulf Stream. Part I: Mooring Design and Instrumentation.

Journal of Atmosheric and Oceanic Technology, 29 (9): 1363–1376. doi:10.1175/JTECH-D-12-00060.1

Yu, L., Jin, X., & Weller, R.A. (2008). *Multidecade Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables*. Woods Hole Oceanographic Institution, OAFlux Project Technical Report. OA-2008-01, 64pp. Woods Hole. Massachusetts.

Zelinka, M., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S., & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letter*, 47, https://doi.org/10.1029/2019GL085782, e2019GL085782, 2020.

Zhang, J., Zuidema, P., Turner, D., & Cadeddu, M. (2018). Surface-based microwave humidity retrievals over the equatorial Indian Ocean: Applications and challenges. *Journal of Applied Meteorology and Climatology*, 57, 1765-1782, doi:10.1175/JAMC-D-17-0301.1

Figures

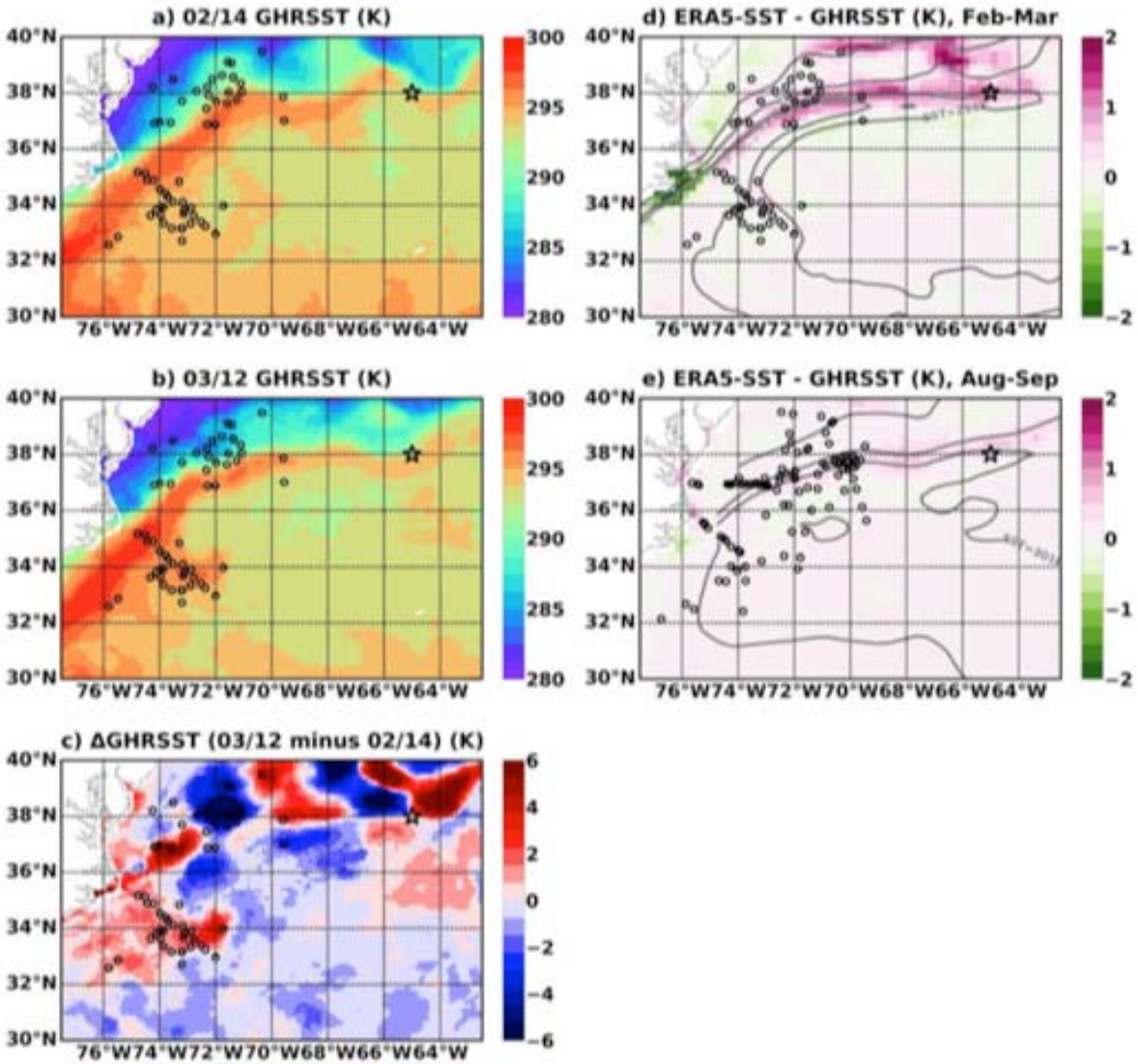


Figure 1. Sea surface temperatures (GHRSST) on a) 14 February 2020, b) 12 March 2020, and c) 12 March - 14 February, 2020. Open circles indicate ACTIVATE dropsonde locations, and the star denotes the CLIMODE buoy location. (ERA5-SST – GHRSST) differences for d) February-March 2020 and e) August-September 2020. Gray contours in d) and e) correspond to ERA5 SSTs of 286, 290, 294 and 295K in panel d) and 301K in panel e).

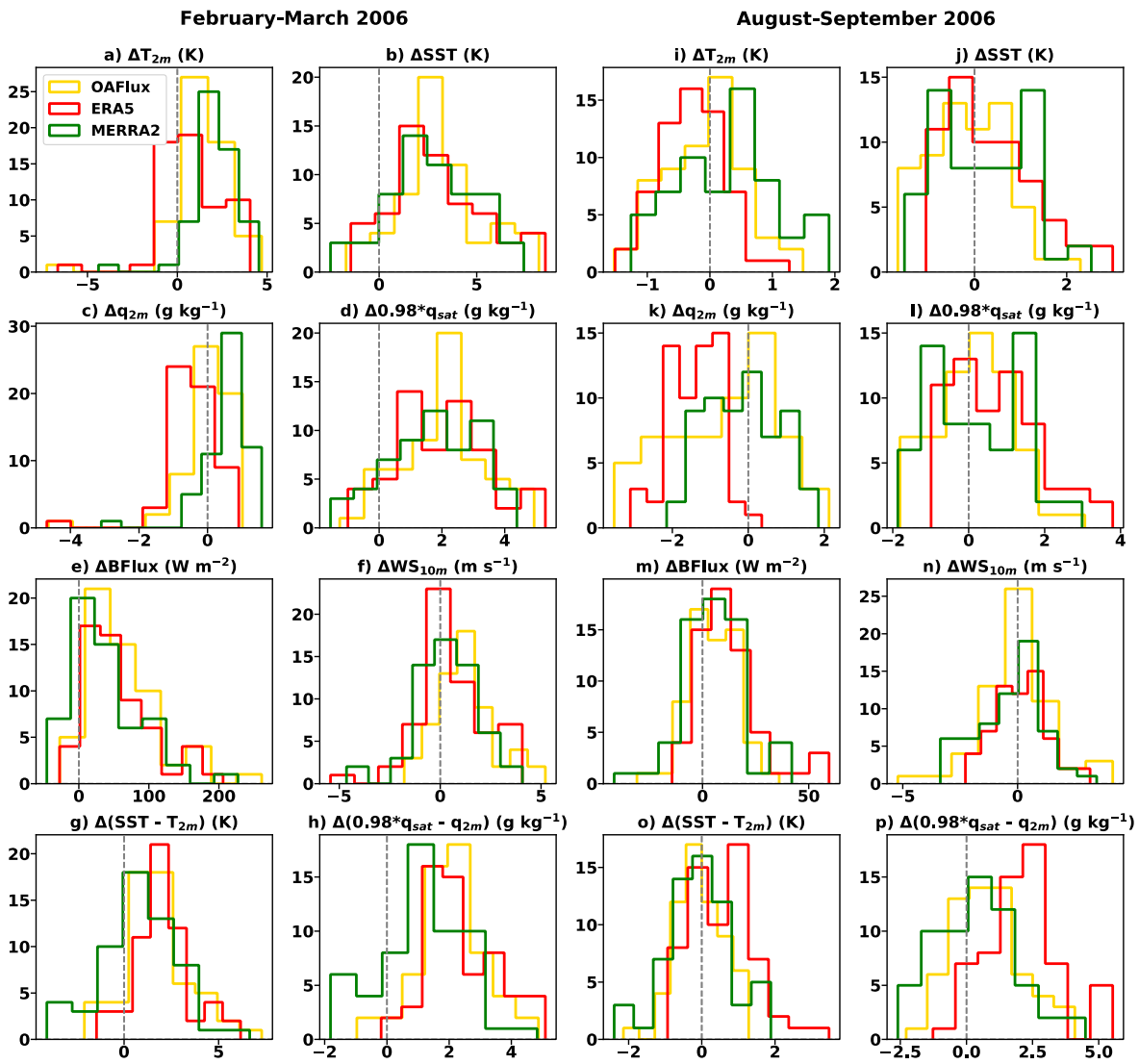


Figure 2. Comparison of ERA5 (red), MERRA2 (green), and OAFLUX (yellow) daily-mean differences from CLIMODE buoy measurements (reanalysis-buoy) for February-March 2006 for a) near-surface air temperature, b) sea surface temperature, c) near-surface specific humidity, d) saturated specific humidity at the SST, e) buoyancy flux, f) wind speed at 10m, and the air-sea g) temperature and h) humidity differences. i)-p): same as a)-h) but for August-September 2006. Note changes in x-scale range. MERRA2 SST is a skin value, while the buoy, ERA5, and OAFLUX SSTs are foundation SST. OAFLUX WS_{10m} is the neutral wind speed.

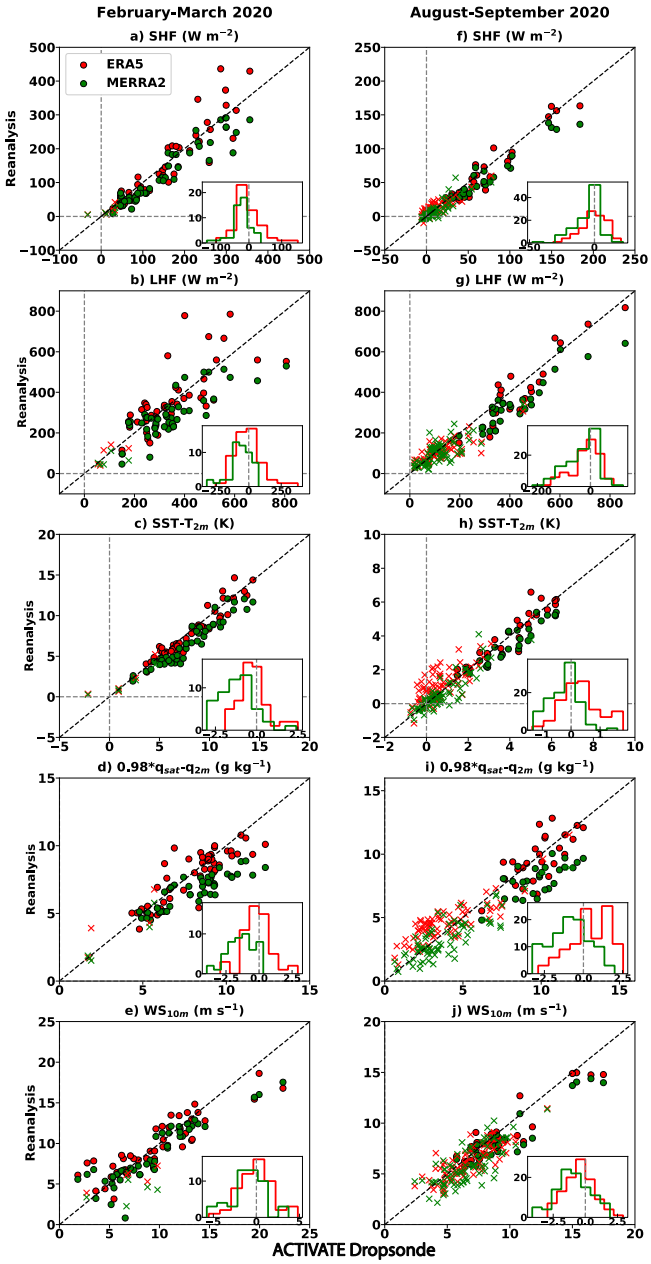


Figure 3. Comparison of nearest-in-space-and-time hourly-mean ERA5 (red) and three-hourly-mean MERRA2 (green) reanalysis to instantaneous ACTIVATE dropsonde-calculated values for February-March 2020 of a) sensible heat flux, b) latent heat flux, calculated using the COARE v3.5 algorithm, c) $SST\text{-}T_{2m}$, d) $0.98*q_s\text{-}q_{2m}$, and e) WS_{10m}. f)-j): same as a)-e) but for August-September 2020; note change in range on both axes. Insets within each panel are histograms of the (reanalysis-dropsonde) differences. Filled circles represent cold-air outbreak conditions, and 'x' markers signify non-CAO conditions, using θ_{SST_skin} - θ_{900hPa} > 0 to define whether an individual dropsonde represented a CAO (see supplement). GHRSST values are corrected to represent a 'skin' SST.

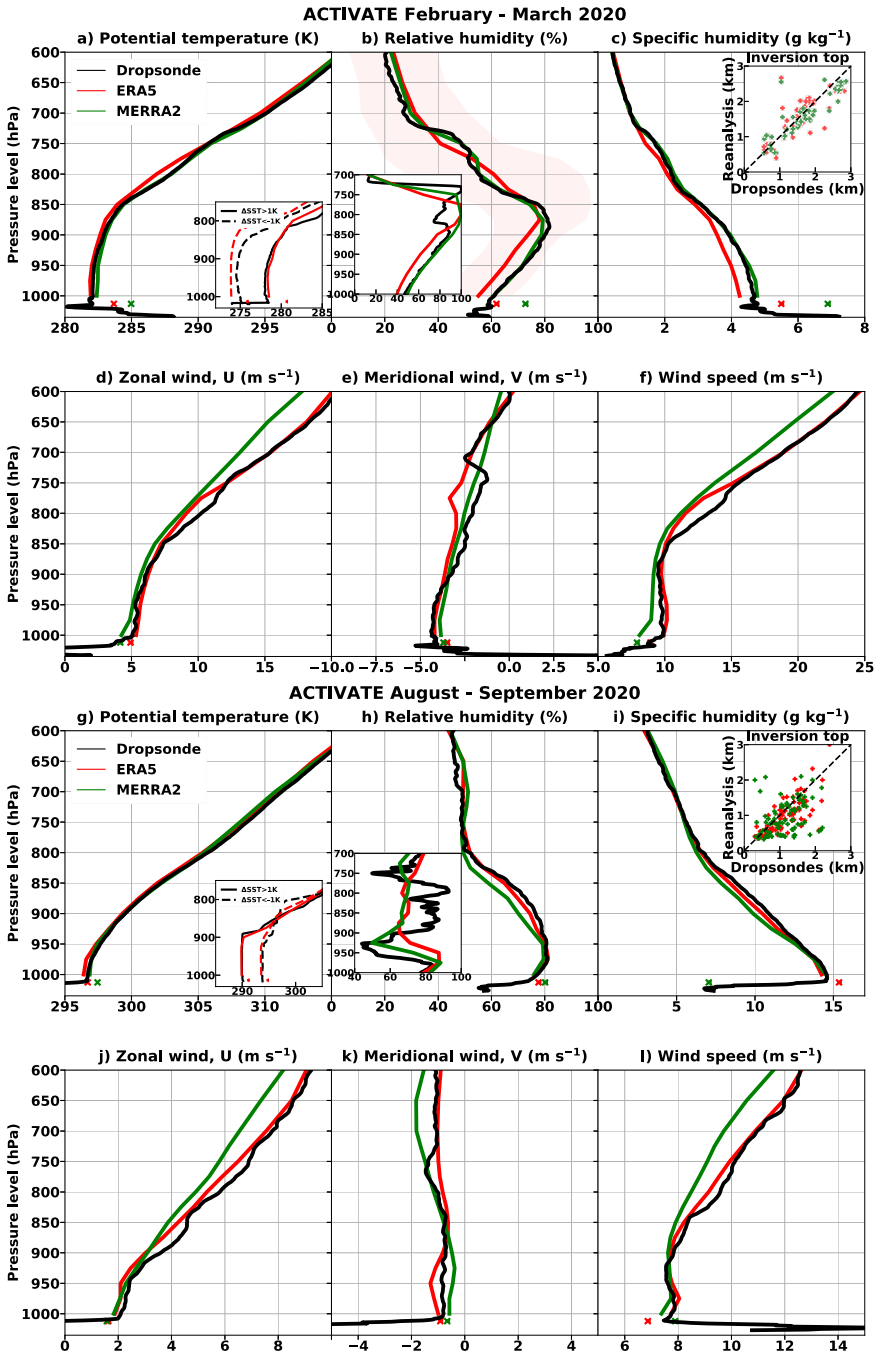


Figure 4. Mean vertical profiles from dropsondes, ERA5, and MERRA2 of a) potential temperature θ, b) relative humidity, c) specific humidity, d) zonal wind (U), e) meridional wind (V), and f) wind speed for February-March 2020. Colored shading indicates the standard deviation. Small 'x' markers indicate θ_{2m}, RH_{2m} and q_{2m}, and U, V and *WS* at 10m. f)-l): same as a)-f) but for August-September 2020. Insets in a) and g) represent the mean θ profiles for ERA5-SST – GHRSST > 1K and ERA5-SST – GHRSST < -1K. The inset profile in panel b) is from 28 February (profile#24) and in panel h) from 20 August (profile#12). Insets in c) and i) indicate inversion top heights estimated from the *RH* profiles.

Figure 1.

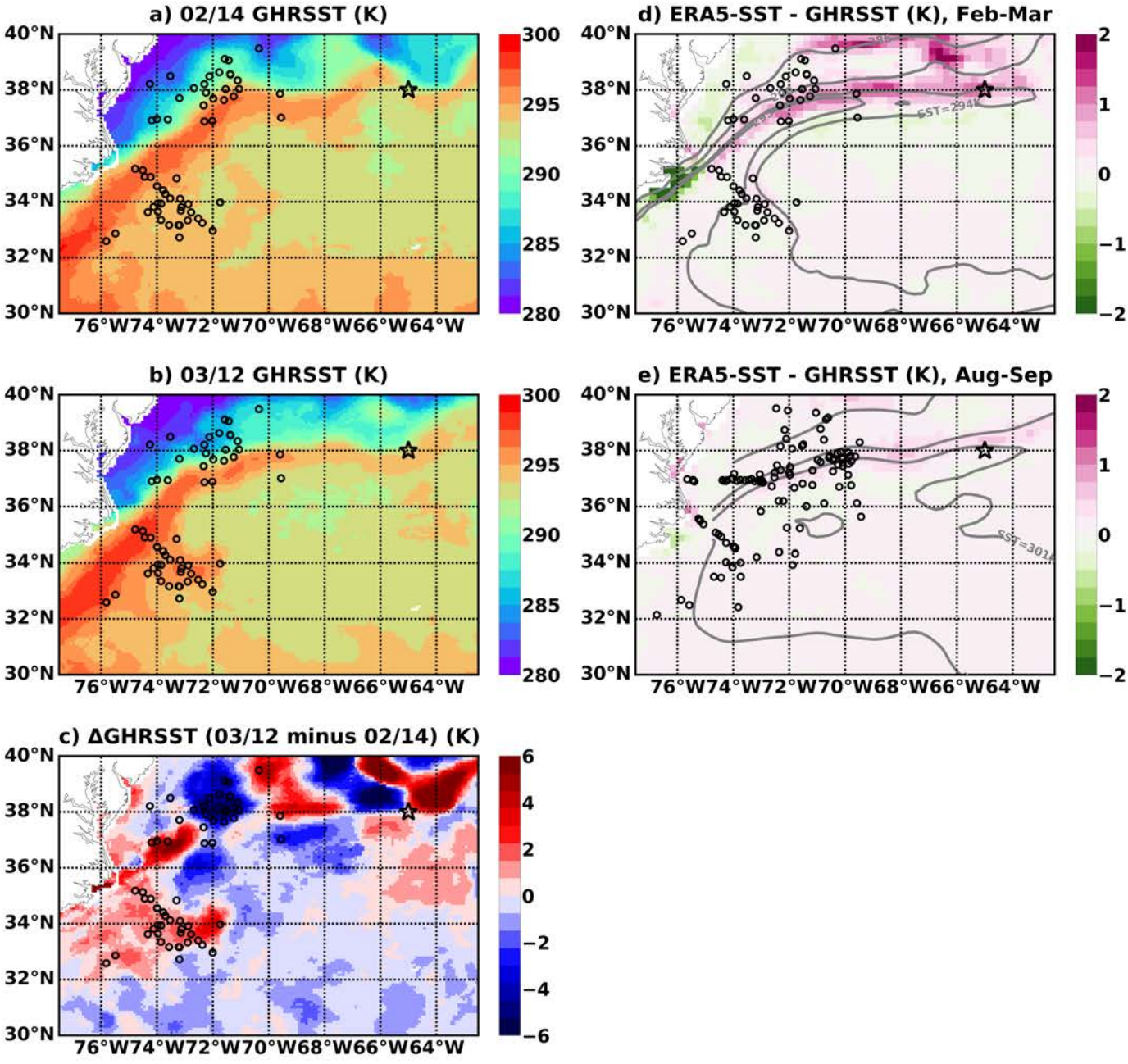

a) 02/14 GHRSST (K)
b) 03/12 GHRSST (K)
c) ΔGHRSST (03/12 minus 02/14) (K)
d) ERA5-SST - GHRSST (K), Feb-Mar
e) ERA5-SST - GHRSST (K), Aug-Sep
SST=294K
SST=301K
40°N
38°N
36°N
34°N
32°N
30°N
76°W74°W72°W70°W68°W66°W64°W
300
295
290
285
280
6
4
2
0
−2
−4
−6
2
1
0
−1
−2

Figure 2.

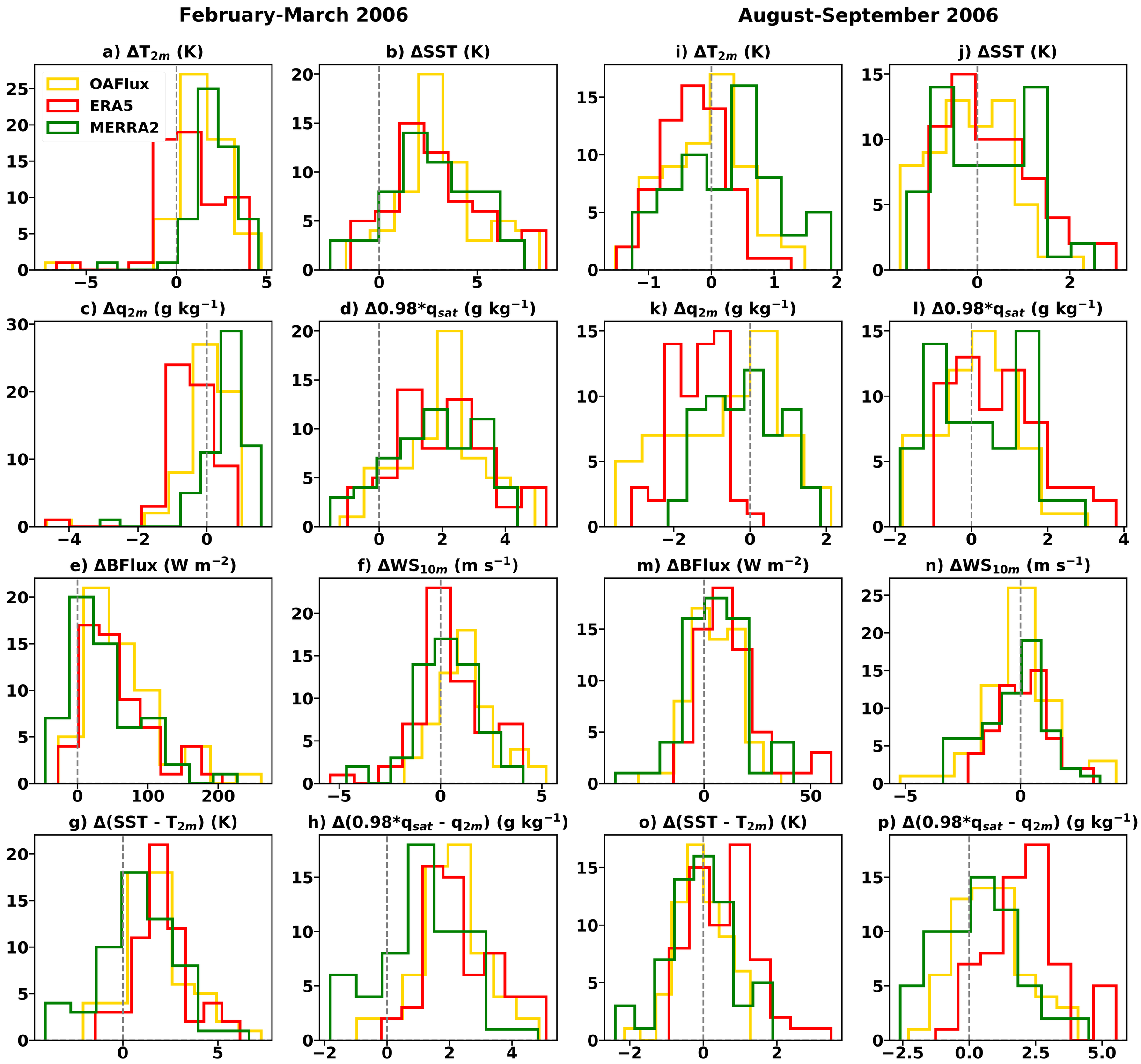

February-March 2006
August-September 2006
a) ΔT2m (K)
b) ΔSST (K)
i) ΔT2m (K)
j) ΔSST (K)
OAFlux
ERA5
MERRA2
c) Δq2m (g kg−1)
d) Δ0.98*qsat (g kg−1)
k) Δq2m (g kg−1)
l) Δ0.98*qsat (g kg−1)
e) ΔBFlux (W m−2)
f) ΔWS10m (m s−1)
m) ΔBFlux (W m−2)
n) ΔWS10m (m s−1)
g) Δ(SST - T2m) (K)
h) Δ(0.98*qsat - q2m) (g kg−1)
o) Δ(SST - T2m) (K)
p) Δ(0.98*qsat - q2m) (g kg−1)

Figure 3.

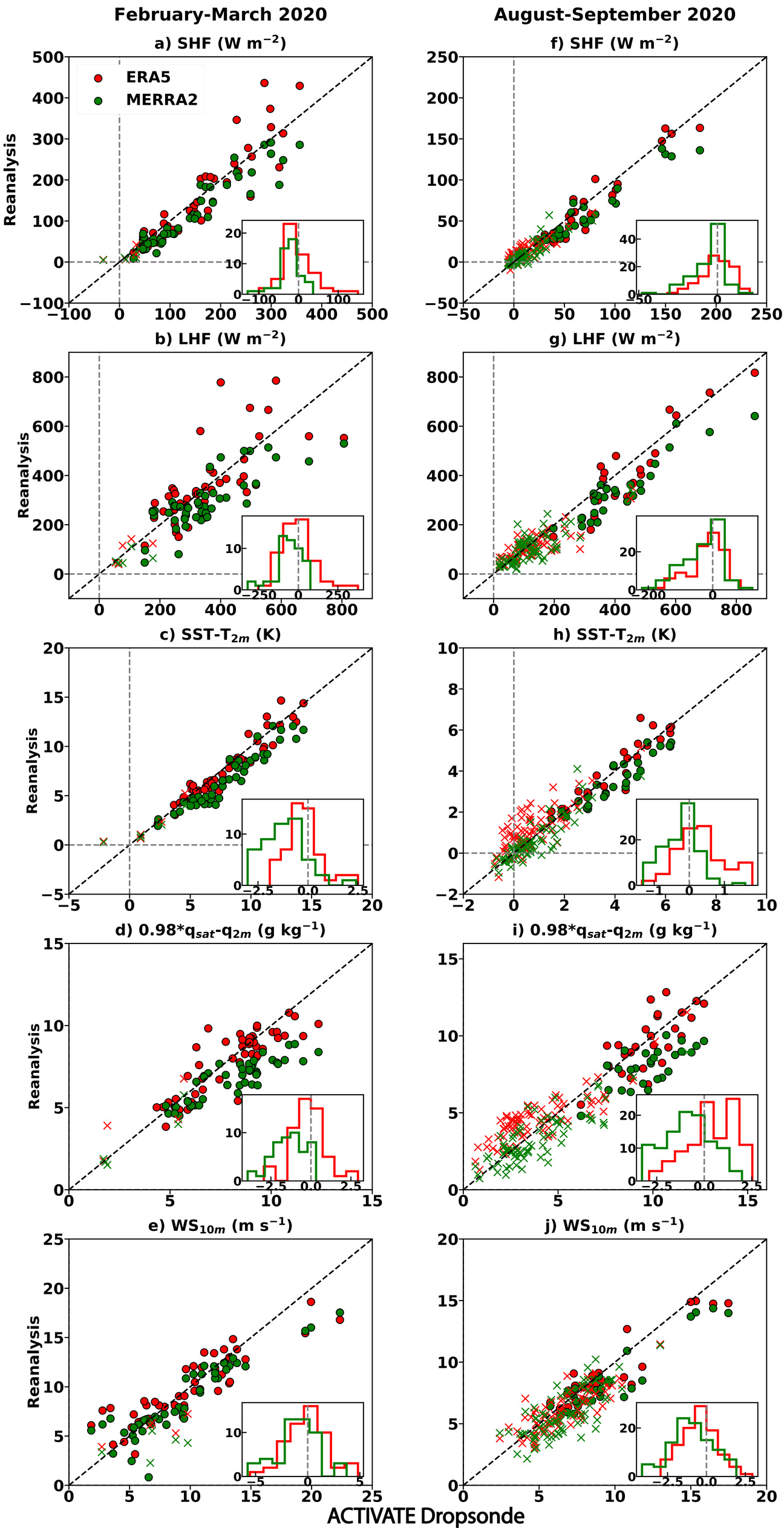
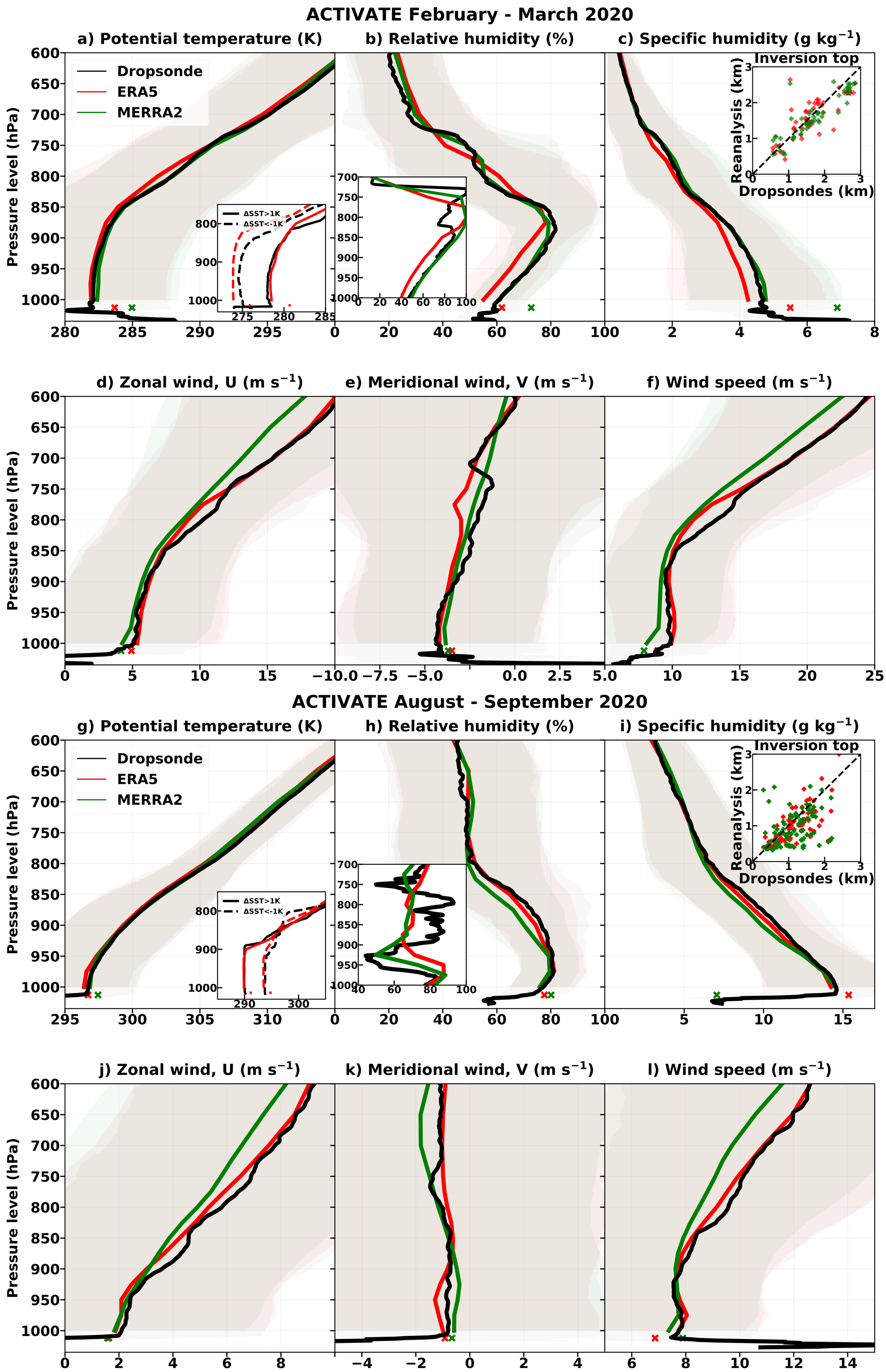

February-March 2020
August-September 2020
a) SHF (W m^{-2})
f) SHF (W m^{-2})
b) LHF (W m^{-2})
g) LHF (W m^{-2})
c) SST-T$_{2m}$ (K)
h) SST-T$_{2m}$ (K)
d) 0.98*q$_{sat}$-q$_{2m}$ (g kg^{-1})
i) 0.98*q$_{sat}$-q$_{2m}$ (g kg^{-1})
e) WS$_{10m}$ (m s^{-1})
j) WS$_{10m}$ (m s^{-1})
ERA5
MERRA2
Reanalysis
ACTIVATE Dropsonde

Figure 4.


ACTIVATE February - March 2020
a) Potential temperature (K)
b) Relative humidity (%)
c) Specific humidity (g kg^{-1})
Dropsonde
ERA5
MERRA2
ΔSST>1K
ΔSST<-1K
Inversion top
Reanalysis (km)
Dropsondes (km)
Pressure level (hPa)
d) Zonal wind, U (m s^{-1})
e) Meridional wind, V (m s^{-1})
f) Wind speed (m s^{-1})
ACTIVATE August - September 2020
g) Potential temperature (K)
h) Relative humidity (%)
i) Specific humidity (g kg^{-1})
j) Zonal wind, U (m s^{-1})
k) Meridional wind, V (m s^{-1})
l) Wind speed (m s^{-1})

Journal of Geophysical Research Letters

Supporting Information for

On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream

Seethala Chellappan[1], Paquita Zuidema[1], Jim Edson[2], Michael Brunke[3], Gao Chen[4], Xiang-Yu Li[5], David Painemal[4,6], Claire Robinson[4], Taylor Shingler[6], Michael Shook[6], Armin Sorooshian[3,7] , Lee Thornhill[6], Florian Tornow[8,9], Hailong Wang[5], Xubin Zeng[3], Luke Ziemba[6]

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL
[2]Woods Hole Oceanographic Institution, Woods Hole, MA
[3]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ
[4]Science Systems and Applications, Inc., Hampton, VA
[5]Pacific Northwest National Laboratory, Richland, WA
[6]NASA Langley Research Center, Hampton, VA
[7]Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ
[8]Earth Institute, Columbia University, New York City, NY
[9]NASA Goddard Institute for Space Sciences, New York City, NY

Contents of this file

Datasets and Method
Tables S1 to S3
Figures S1 to S5

35
36
37

Datasets and Method:

The CLIMODE mooring sampled along the northern periphery of the Gulf Stream in late January-mid February 2006 and again in late March-April 2006, and was otherwise within warmer waters (Weller et al., 2012), with its position relative to the Gulf Stream shown using GHRSST in Fig. S1. The anchor line maintains the buoy within 5-7 km of its nominal position. The Air-Sea Interaction Meteorological system made continuous measurements of temperature (*T*), relative humidity (*RH*) and wind speed (*WS*) at about 3 m above the waterline and measured the SST at a depth of 0.89 m. Radiometers provided downwelling solar and IR radiative fluxes used to estimate the 'skin' sea surface temperature (Fairall et al., 1996). A Direct Covariance Flux System allowed direct covariance computations of buoyancy fluxes (and surface wind stresses; Edson et al., 2013).

GHRSST optimally integrates cloud-penetrating microwave SST data with infrared data of a higher spatial resolution into a daily global SST dataset. We use a 9 km v5.0 MW_IR OISST product available from Remote Sensing Systems.

The reference atmospheric structure information comes from precalibrated NCAR (National Centre for Atmospheric Research) Dropsonde 94 (NRD94s; Wick et al., 2018), released from the NASA Langley Beechcraft UC12 research aircraft at an approximate flight altitude of 9 km. Pressure, temperature and humidity data are returned at 2 Hz, and of winds at 4 Hz, corresponding to a vertical resolution of 6-15 m. The pressure, temperature, humidity and wind speed are resolved to 0.1 hPa, 0.1^0C, 1% and 0.1 m s^{-1}, respectively, with a standard deviation of differences between two successive repeated calibrations of 0.4 hPa, 0.2^0C, 2% and 0.2 m s^{-1}, respectively.

ERA5 is the fifth-generation global atmospheric reanalysis developed by the European Centre of Medium-range Weather Forecast (ECMWF), described comprehensively within Hersbach et al., (2020). ERA5 relies on a 12-hr 4D-var Integrated Forecasting System (IFS) cycle 41R2 data assimilation, with data available every hour at a horizontal resolution of 31 km, gridded to 0.25^0. The ERA5 atmospheric model is coupled with a land surface and a wave model, with internal computations encompassing 137 vertical levels, of which 37 are output, including the lowest level at 10 m. ERA5 estimates of the temperature and specific humidity at 2 m altitude (T_{2m} and q_{2m}), developed using Monin-Obukhov theory to relate the skin sea surface temperature and its saturated specific humidity to the 10 m model level,

support explicit comparisons to buoy measurements. Rivas and Stoffelen, (2019) document improved mid-latitude storm track surface wind representations compared to the previous ERA-Interim, attributed to a higher vertical resolution (137 vs 60 model levels).

MERRA2 (Modern Era Retrospective analysis for Research and Applications; Bosilovich et al., 2015) relies on the Goddard Earth Observing System (GEOS-5.12) atmospheric global model (Molod et al., 2015) combined with the Gridpoint Statistical Interpolation data assimilation system (Wu et al. 2002). MERRA2 assimilates microwave and infrared radiances and select retrievals from polar-orbiters and geostationary satellites, including aerosol optical depth. Daily 0.25^{o}x0.25^{o} Reynolds SSTs (Reynolds et al., 2007) were prescribed until March 2006, and thereafter, high resolution satellite-derived daily SSTs similar to the OISSTs used by ERA5. Recent relevant improvements include an improved relationship between the ocean surface roughness and ocean surface stress (Molod et al., 2015). The change from MERRA to MERRA2 reduces surface wind speeds in the ACTIVATE region, by approximately 2 m s^{-1}, but surface turbulent fluxes are marginally affected, and a change in the critical relative humidity for cloud condensation also produce little change in the boundary layer q, both just for the ACTIVATE region (Molod et al., 2015).

The bulk fluxes are calculated from the dropsondes using the TOGA-COARE v.3.5 bulk flux algorithm. This relies on Monin-Obukhov similarity theory as:

$$Q_H = \rho\, C_H\, U \left(S_{surf} - S_{2m}\right) \qquad \text{---------- (S1)}$$

$$Q_E = \rho\, L_v\, C_E\, U\, (q_{surf} - q_{2m}) \qquad \text{---------- (S2)}$$

$$Q_B = Q_H\, (1 + 0.6\, q_{2m}) + Q_E\, 0.61\, \frac{C_p}{L_v} T_{2m} \qquad \text{---------- (S3)}$$

where Q_H, Q_E and Q_B surface sensible, latent and buoyancy heat fluxes respectively and S is the dry static energy. S_{surf} is based on the (foundation) GHRSST corrected to be the surface 'skin' SST value, and q_{surf} is 98% of the surface saturation humidity (q_{sat}) accounting for the salinity effect. ρ is air density, L_v is the latent heat of vaporization, C_p is the specific heat at constant pressure, U the wind speed accounting the free convection velocity w*, and, C_H and C_E the bulk transfer coefficient for heat and moisture, respectively.

Cold-air outbreak conditions were determined in a separate manner for Table S1 and Figure 3. Table S1 classifies an entire flight as either a CAO or a non-CAO flight, based on a visual identification using MODIS imagery from the NASA Worldview URL site (worldview.earthdata.nasa.gov). Days with obscuring cirrus or a lack of low clouds were excluded. For the individual dropsonde analysis shown in Fig. 3, dropsondes were classified as 'CAO' or 'non-CAO' using the potential difference between the estimated 'skin GHRSST_skin and that at 900 hPa. This definition emphasizes the surface forcing contribution more than that of the cloud capping inversion than the 850 hPa level applied within Papritz et al., 2015. The lower level was chosen as it clearly avoids falling within the stratiform cloud layer for the deeper boundary layers of the ACTIVATE domain (e.g., Fig. 4b, inset).

Bosilovich, M. G. and Coauthors. (2015). *MERRA-2: Initial evaluation of the climate. Technical Report Series on Global Modeling and Data Assimilation*, Vol. 43, NASA Tech. Rep. NASA/TM– 2015–104606, 139 pp. [Available online at https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf.]

Edson, J. B., Jampana, V., Weller, R.A., Bigorre, S.P., Plueddemann, A.J., Fairall, C.W., Miller, S.D., Mahrt, L., Vickers, D., & Hersbach, H. (2013). On the exchange of momentum over the open ocean. *Journal of Physical Oceanography*, 43, 1589–1610. doi:10.1175/ JPO-D-12-0173.1.

Fairall, C.W., Bradley, E.F., Godfrey, J.S. , Edson, J.B., Young, G.S. and Wick, G.A. (1996) Cool skin and warm layer effects on the sea surface temperature. *J. Geophys. Res., 101,* 1295-1308.

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global analysis. *Quarterly Journal of Royal Meteorological Society*, 146, 1999-2049. doi:10.1002/qj.3803.

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development,* 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Papritz, L., Pfahl, S., Sodemann, H., & Wernli, H. (2015). A Climatology of Cold Air Outbreaks and Their Impact on Air–Sea Heat Fluxes in the High-Latitude South Pacific, *Journal of Climate*, *28*, 342-364. doi:10.1175/JCLI-D-14-00482.1

Reynolds, R.W. & Smith, T.M. (1994). Improved Global Sea Surface Temperature Analysis using Optimum Interpolation. *Journal of Climate*, 7, 929-948. doi: 10.1175/1520-0442(1994)007

Rivas, M. B., & Stoffelen, A. (2019). Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Science*, 15, 831-851. doi: 10.5194/os-15-831-2019

Weller, R.A., Bigorre, S.P., Lord, J., Ware, J.D., & Edson, J.B. (2012). A Surface Mooring for Air–Sea Interaction Research in the Gulf Stream. Part I: Mooring Design and Instrumentation. *Journal of Atmosheric and Oceanic Technology*, 29 (9): 1363–1376. doi:10.1175/JTECH-D-12-00060.1

Wick, G. A., Hock, T.F., Neiman, P.J., Vömel, H., Black, M.L., & Spackman, J.R. (2018). The NCAR–NOAA Global Hawk Dropsonde System. *Journal of Atmospheric and Oceanic Technolology*, 35, 1585–1604. doi:10.1175/JTECH-D-17-0225.1.

Wu, W.-S., Purser, R. J., & Parrish, D. F. (2002). Three-dimensional variational analysis with spatially inhomogeneous covariances. *Monthly Weather Review*, 130, 2905–2916, doi:10.1175/ 1520-0493(2002)130,2905:TDVAWS.2.0.CO;2.

Winter 2020 (mm-dd)	Cloud type (#profiles)	Summer 2020 (mm-dd)	Cloud type (#profiles)
02-14	Cirrus-obscured (4; CAO*)	08-13	non-CAO (5)
02-15	CAO (4)	08-17	non-CAO (6)
02-17	clear (4; non-CAO*)	08-20	non-CAO (5)
02-27	CAO (2)	08-21	non-CAO (5)
02-28	CAO (13)	08-25	non-CAO (6)
02-29	CAO (2)	08-26	non-CAO (6)
03-01	CAO (13)	08-28	non-CAO (8)
03-02	CAO (2)	09-02	non-CAO (6)
03-06	CAO (3)	09-03	non-CAO (6)
03-08	CAO (4)	09-10	non-CAO (4)
03-09	non-CAO (2)	09-11	non-CAO (6)
03-11	non-CAO (2)	09-15	CAO (6)
03-12	non-CAO (4)	09-16	no dropsondes
		09-21	CAO (5)
		09-22	CAO (7)
		09-23	clear (8; non-CAO*)
		09-29	non-CAO (13)
		09-30	non-CAO (5)

*CAO under cirrus or non-CAO, using definition of Papritz et al., (2015); $\theta_{SKT} - \theta_{850} > 0$ for CAO.

Table S1. Winter and summer 2020 ACTIVATE flight days, cloud conditions and dropsonde number. Eight of the 13 UC-12 King Air flight days in February-March, 2020 sampled cloudy cold-air outbreak (CAO) conditions, encompassing 43 of 59 dropsondes total and 3 of the 18 flight days in August-September 2020, encompassing 18 of 107 dropsondes total). Cloudy CAO conditions were visually determined from satellite imagery and excluded days with obscuring cirrus.

CLIMODE		ERA5 - CLIMODE		MERRA2 - CLIMODE		OAFLUX - CLIMODE	
		Feb-Mar	*Aug-Sep*	*Feb-Mar*	*Aug-Sep*	*Feb-Mar*	*Aug-Sep*
BFlux	*Bias*	55	11	39	5	63	5
	RMS	49	15	50	14	54	11
	Correlation	0.85	0.88	0.84	0.86	0.84	0.9
SST-T_{2m}	*Bias*	2.1	0.6	0.7	-0.08	1.8	-0.04
	RMS	1.5	0.8	2	0.9	1.7	0.7
	Correlation	0.9	0.85	0.83	0.83	0.88	0.91
$0.98*q_{sat}$-q_{2m}	*Bias*	2.4	2.1	1.2	0.4	2.1	0.8
	RMS	1.2	1.4	1.4	1.5	1.1	1.3
	Correlation	0.83	0.90	0.77	0.89	0.84	0.93
WS_{10m}	*Bias*	0.5	0.13	0.3	-0.35	1.2	-0.03
	RMS	1.7	1.1	1.5	1.4	1.4	1.5
	Correlation	0.90	0.94	0.92	0.91	0.94	0.88

Table S2. The mean bias, root mean square (RMS) deviation, and correlation between daily-mean CLIMODE buoy and ERA5, MERRA2 and OAFLUX values of buoyancy flux (W m^{-2}), SST-T_{2m} (K), $0.98*q_{sat}$-q_{2m} (g kg^{-1}), and WS_{10m} (m s^{-1}) depicted for February-March and August-September in 2006.

ACTIVATE		ERA5 - ACTIVATE		MERRA2 - ACTIVATE	
		Feb-Mar	*Aug-Sep*	*Feb-Mar*	*Aug-Sep*
SHF	*Bias*	-2.5	-0.24	-22.8	-4.84
	RMS	41	9.62	30	10.12
	Correlation	0.93	0.97	0.95	0.97
LHF	*Bias*	-4	-19.61	-58	-31.12
	RMS	106.5	54.88	86.5	58.67
	Correlation	0.81	0.94	0.87	0.95
SST-T_{2m}	*Bias*	-0.2	0.32	-1.15	-0.26
	RMS	0.91	0.66	1.03	0.50
	Correlation	0.97	0.93	0.96	0.96
$0.98*q_{sat}$-q_{2m}	*Bias*	-0.16	0.45	-1.35	-0.87
	RMS	1.06	1.25	1.05	1.23
	Correlation	0.89	0.93	0.93	0.94
WS_{10m}	*Bias*	-0.12	-0.47	-0.97	-0.92
	RMS	1.99	1.17	2.02	1.31
	Correlation	0.89	0.9	0.88	0.87
Inversion top height	*Bias*	-74	-205	-96	-186
	RMS	-375	-425	-323	-587
	Correlation	0.83	0.62	0.89	0.48

Table S3. The mean bias, root mean square (RMS) deviation, and correlation between ACTIVATE dropsonde and ERA5/MERRA2 values of sensible and latent heat fluxes (SHF and LHF respectively, W m^{-2}), SST-T_{2m} (K), $0.98*q_{sat}$-q_{2m} (g kg^{-1}), WS_{10m} (m s^{-1}), and the inversion top height (m), estimated using a relative humidity threshold for February-March and August-September in 2020.

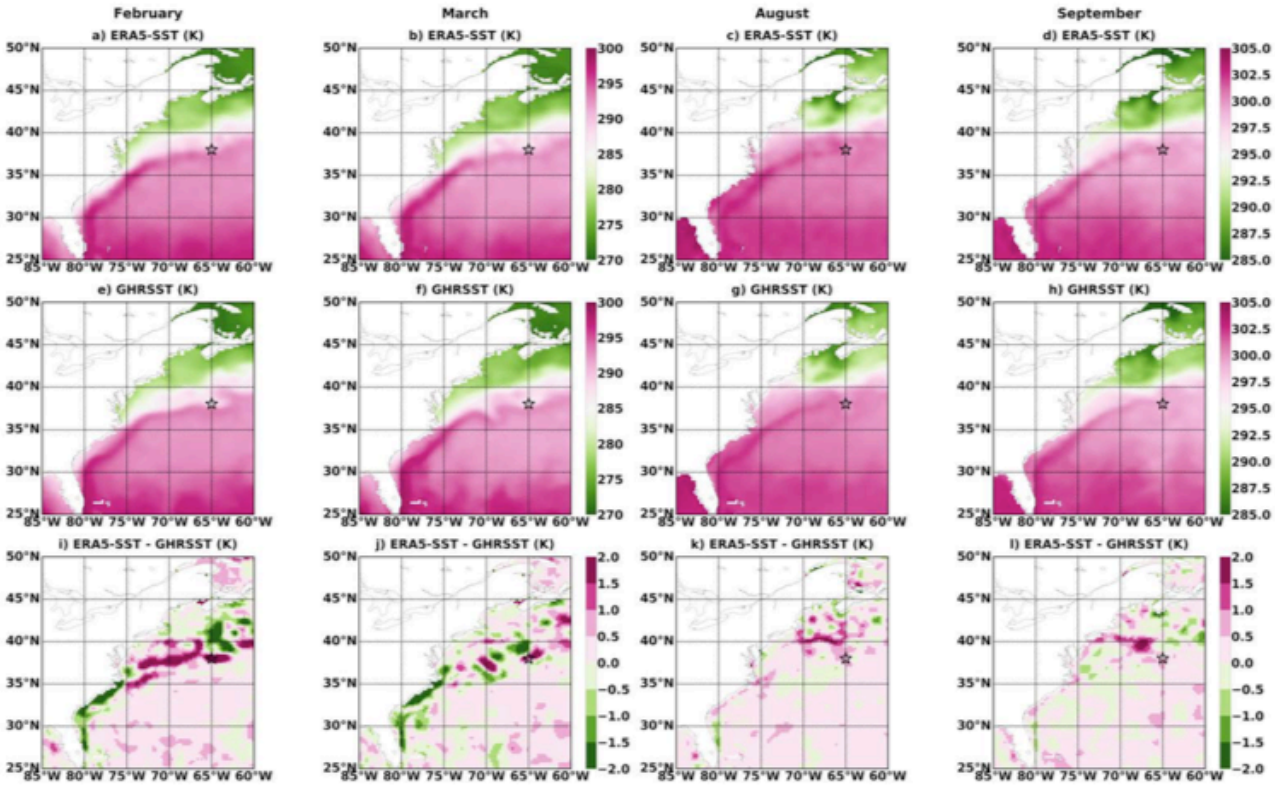


Figure S1. Top row: ERA5 foundation SST spatial distribution during February, March, August, and September in 2006. Middle row: same as top but for GHRSST. Bottom row: (ERA5-SST – GHRSST) difference for the same four months.

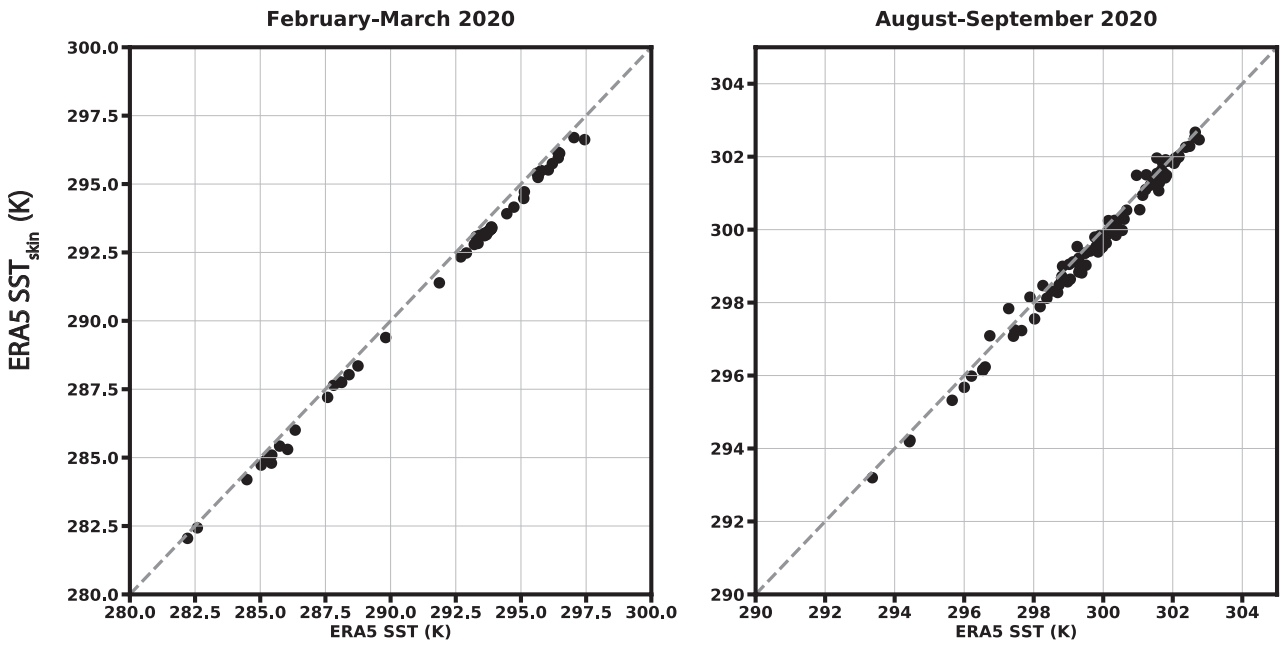


Figure S2. ERA5 skin sea surface temperature (SST_{skin}) as a function of its foundation SST at 1-m depth, at dropsonde locations and times for left) February-March 2020 and right) August-September 2020.

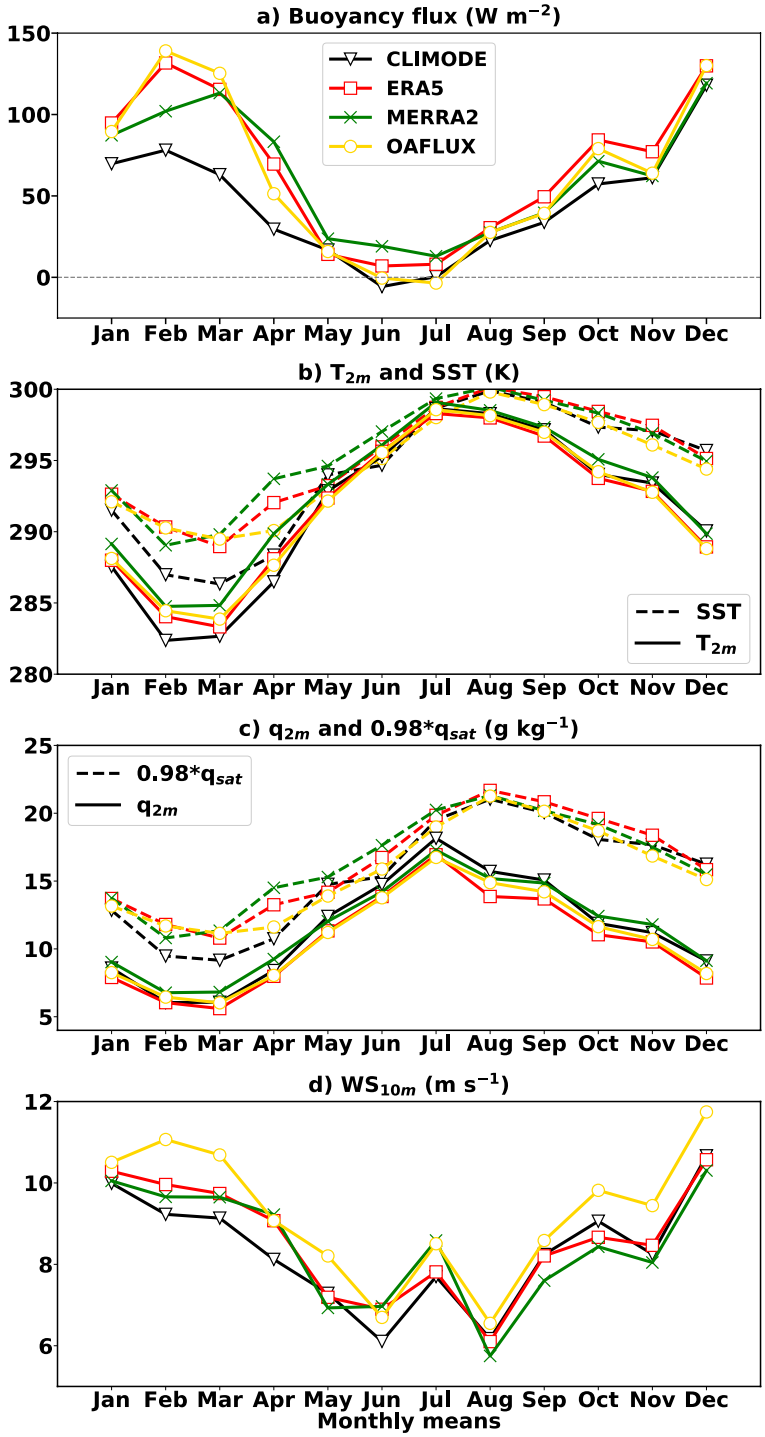


Figure S3. Monthly-means at the CLIMODE buoy in year 2006 of a) buoyancy flux, b) SST and T_{2m}, c) 0.98*q_s and q_{2m}, and d) 10 m wind speed (WS_{10m}) for the CLIMODE buoy (black), ERA5 (red), MERRA2 (green), and OAFLUX (yellow). MERRA2 SST is a skin value, while the buoy, ERA5, and OAFLUX SSTs are foundation SSTs. OAFLUX WS_{10m} is the neutral wind speed.

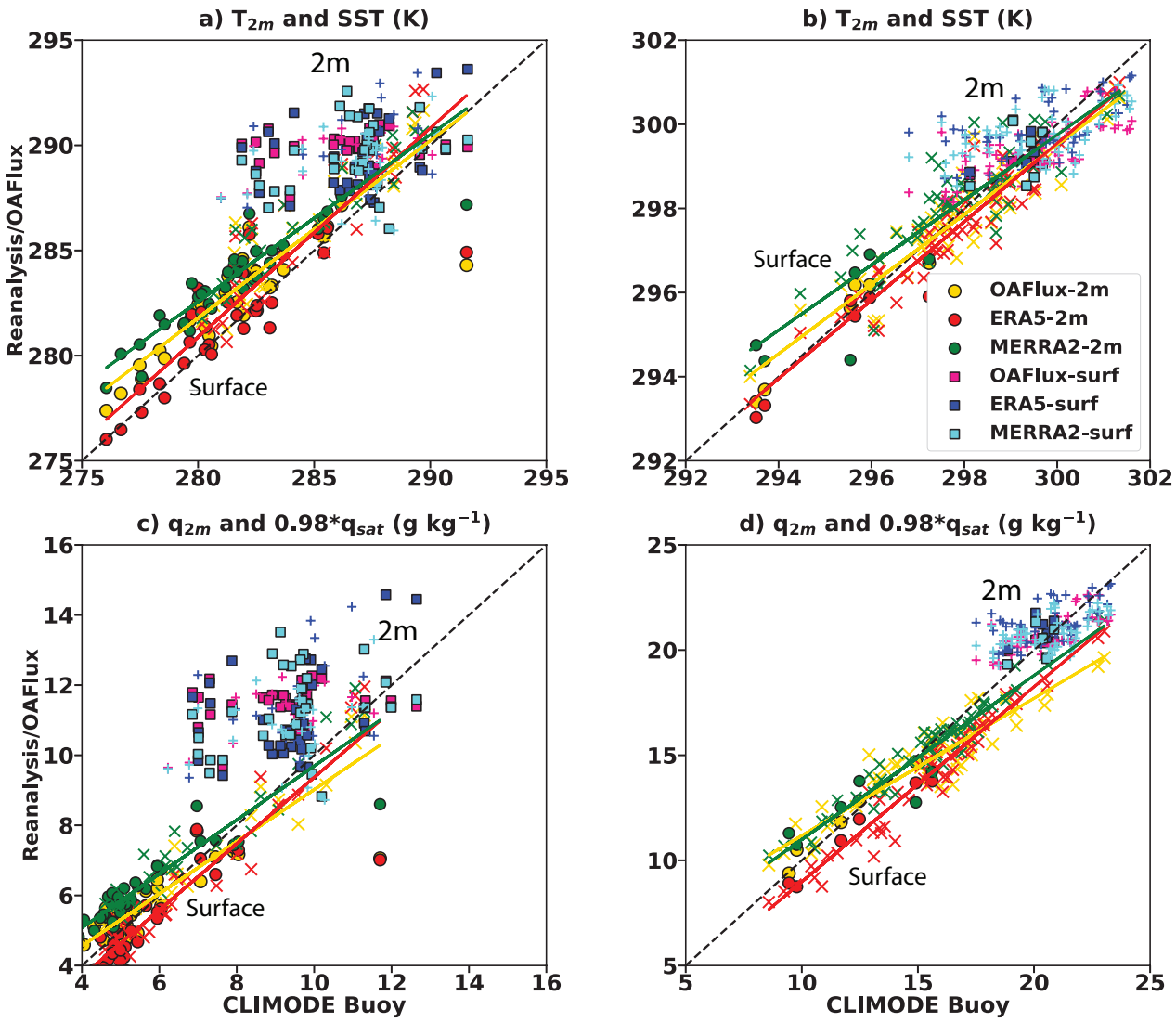


Figure S4. Comparison of ERA5, MERRA2 and OAFLUX daily mean surface meteorology against left) CLIMODE buoy for February-March 2006 a) SST and T_{2m}, and c) q_s and q_{2m}. Right two panels are the same as the left panels but for August-September 2006. The filled circles represent the days with cold-air outbreak conditions, whereas 'x' denotes non-CAO days.

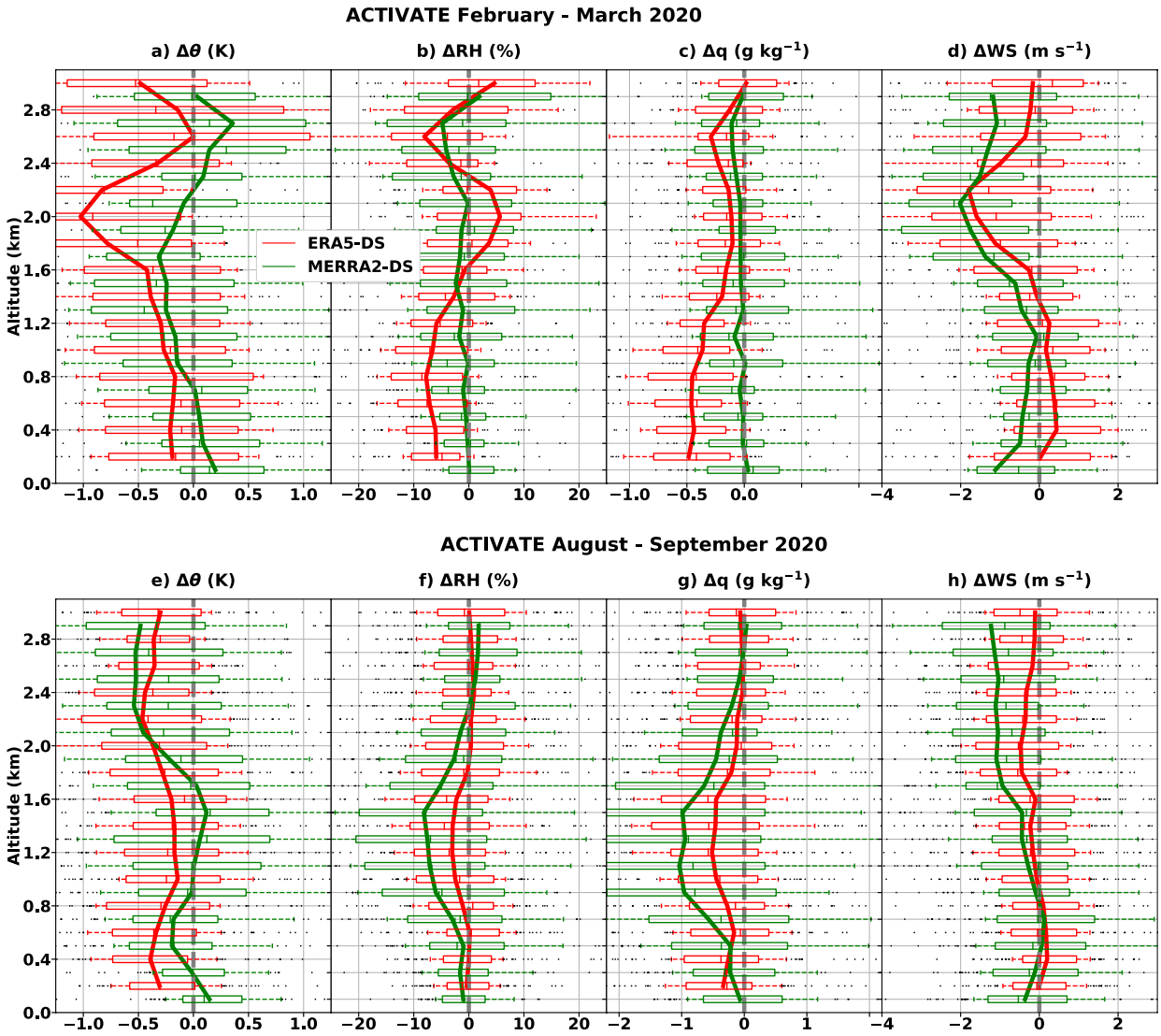


Figure S5. Mean difference between dropsondes and reanalysis profiles (ERA5 in red and MERRA2 in green; reanalysis-dropsonde) in a) potential temperature, b) relative humidity, c) specific humidity, and d) wind speed of February-March 2020, shown as the interquartile range (horizontal bars), 15–85 percentile (thin horizontal dashed line), and median (thin vertical line). e)-h): same as a)-d) but for August-September 2020 deployment.

On assessing ERA5 and MERRA2 representations of cold-air outbreaks across the Gulf Stream

Seethala Chellappan[1], Paquita Zuidema[1], Jim Edson[2], Michael Brunke[3], Gao Chen[4], Xiang-Yu Li[5], David Painemal[4,6], Claire Robinson[4], Taylor Shingler[6], Michael Shook[6], Armin Sorooshian[3,7], Lee Thornhill[6], Florian Tornow[8,9], Hailong Wang[5], Xubin Zeng[3], Luke Ziemba[6]

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, FL
[2]Woods Hole Oceanographic Institution, Woods Hole, MA
[3]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, AZ
[4]Science Systems and Applications, Inc., Hampton, VA
[5]Pacific Northwest National Laboratory, Richland, WA
[6]NASA Langley Research Center, Hampton, VA
[7]Department of Chemical and Environmental Engineering, University of Arizona, Tucson, AZ
[8]Earth Institute, Columbia University, New York City, NY
[9]NASA Goddard Institute for Space Sciences, New York City, NY

Corresponding authors: Seethala Chellappan (seethala.chellappan@rsmas.miami.edu) and Paquita Zuidema (pzuidema@miami.edu)

Key Points:

- Reanalysis surface fluxes and boundary layers are representative of observations to first-order, sufficient for higher-resolution model initialization.
- Reanalyses represent the Gulf Stream more broadly than is seen in nature, contributing to turbulent flux and boundary layer biases.
- Previously-noted thermodynamic and dynamic biases reinforce (ERA5) or compensate (MERRA2) surface fluxes but support realistic winter boundary layer heights.

Abstract

The warm Gulf Stream sea surface temperatures (SSTs) strongly impact the evolution of winter clouds behind atmospheric cold fronts. Such cloud evolution remains challenging to model. The Gulf Stream is too wide within the ERA5 and MERRA2 reanalyses, affecting the turbulent surface fluxes. Known problems within the ERA5 boundary layer (too-dry and too-cool with too strong westerlies), ascertained primarily from ACTIVATE 2020 campaign aircraft dropsondes and secondarily from older buoy measurements, reinforce surface flux biases. In contrast, MERRA2 winter surface winds and air-sea temperature/humidity differences are slightly too weak, producing surface fluxes that are too low. Reanalyses boundary layer heights in the strongly-forced winter cold-air-outbreak regime are realistic, whereas late-summer quiescent stable boundary layers are too shallow. Nevertheless, the reanalysis biases are small, and reanalyses adequately support their use for initializing higher-resolution cloud process modeling studies of cold-air outbreaks.

Plain Language Summary

The Gulf Stream is a narrow band of warm water to the east of continental north America. As air moves eastward off of the continent, the warm ocean temperatures transfer moisture and heat that help develop and modify the marine low clouds. This transfer, particularly during cold-air outbreaks present significant modeling challenges that contribute uncertainty to temperature projections for a world with more carbon dioxide. Simulations seeking to represent the details of such shallow clouds must rely on initializations and forcings that originate from coarser-resolution renalyses. Here we explore how well two major reanalyses and a commonly-used flux product represent these fluxes and the boundary layer, using ocean buoy measurements and new *in-situ* observations from an aircraft campaign. We find that the reanalyses are adequate for the purpose of initializing higher-resolution modeling of the cold-air outbreak clouds. In particular, the winter boundary layer heights are realistic. These heights are important for capturing winter cloud-environmental interactions correctly. Late summer boundary layers are too shallow. Known biases do remain present, and impact the surface flux errors differently in the two reanalyses examined.

1. Introduction

A prominent feature of the northwest Atlantic is the Gulf Stream, a western boundary current transporting warm waters to the north. The Gulf Stream is up to 10°C warmer than the surrounding waters (Fig. 1), fueling atmospheric convection. During the off-summer months, westward-moving mid-latitude synoptic disturbances exchange warm, low-latitude air with colder air. In this situation, the air-sea interaction becomes exceptional. The evolution of the low marine clouds in the post-frontal regions of strong subsidence is described in Grossman and Betts (1990); Kolstad et al., (2009); Liu et al., (2014); Fletcher et al., (2016a, 2016b); McCoy et al., (2017), and Painemal et al., (2021). These clouds remain challenging to model (Skyllingstad and Edson, 2009; Field et al., 2014; Abel et al., 2017). Leading questions remain realistic representations of the roll cloud circulations (Honnert et al, 2020), and the correct partitioning between the liquid and ice phases (Mulmenstadt et al., 2014; Field and Heymsfield, 2014). The latter contributes to the cloud feedback uncertainty in climate models (Zelinka et al., 2020; Sherwood et al., 2020).

The sharp increase in the sea surface temperature (SST) of the Gulf Stream is particularly noticeable in winter. Turbulent fluxes can exceed 1000 W m^{-2} during cold-air outbreak events (Bane and Osgood, 1989; Marshall et al., 2009; Biggore et al., 2013). The adjustment of the boundary-layer air temperature and humidity to the underlying surface can establish a thermally-direct circulation, in which horizontal pressure and boundary-layer height gradients drive a surface wind convergence on the warmer flank of the Gulf Stream (Minobe et al., 2008; Liu et al., 2014; Plagge et al., 2016). The increase in surface winds is aided by a downward transfer of momentum and the shear mediates an adjustment to the altered boundary layer stratification as well (Small et al., 2008).

The northwest Atlantic is a strategic environment for improving the understanding and modeling of cold-air outbreaks (CAOs) through observations, aided by the proximity to the eastern north American seaboard. The characterization and process modeling of CAOs is a key objective of the NASA Earth Venture Suborbital-3 Aerosol-Cloud-meTeorology Interactions oVer the western Atlantic Experiment (ACTIVATE; Sorooshian et al., 2019). The process modeling activities rely on reanalysis data for initialization (Tornow et al., 2021; Li et al., 2021), and include subsequent nudging to above-inversion values (Tornow et al., 2021). The ACTIVATE winter 2020 campaign included two dropsonde circles to explicitly derive vertical

velocities following Bony and Stevens, (2019), out of concern that reanalysis-derived vertical motion might not be adequate for model forcings. Reanalysis fluxes provide a tempting alternative to derived *in-situ* fluxes requiring long, low-altitude level legs that compete with the gathering of new cloud microphysical information. Reanalyses, in combination with satellite datasets, also provide useful longer-term context (Painemal et al., 2021).

This study addresses the following two questions: 1) How accurate are the surface fluxes from the latest major reanalyses (the fifth-generation ECMWF (ERA5) and Modern-Era Retrospective Analysis for Research and Applications-2 (MERRA2)) and the Objectively Analyzed air-sea Heat Fluxes (OAFLUX) in the presence of wintertime CAOs over the Gulf Stream? This extends prior assessments based on coarser-resolution products (Moore and Renfrew, 2002; Jin and Yu, 2013. 2) Can ERA5 and MERRA2 provide a realistic depiction of the Gulf Stream-affected boundary layer? We rely on buoy and ACTIVATE dropsonde data for reference. The questions are relevant beyond the scope of the ACTIVATE campaign, and recognize the challenges inherent to representing strong air-sea coupling events. Reanalysis products can be an alternative to observations in weather and climate studies and are also applied to climate model assessments – sometimes without fully understanding their performance.

2. Datasets and Method

Buoy measurements including direct covariance buoyancy fluxes from the CLIMODE (Climate Variability and Predictability Mode Water Dynamics Experiment; Marshall et al., 2009) campaign provide absolute reference values. The CLIMODE buoy was situated at 38°N, 65°W (Fig. 1), with the data from the 12 months of year 2006 incorporated into this study. The buoy was to the north of the Gulf Stream in Febuary-March, 2006, and in August-September, 2006, within and south of the Gulf Stream (Fig. S1). Twenty-minute averages of temperature, relative humidity and wind speed were measured at about 3m above the waterline and calibrated, with drifts and biases corrected (Weller et al., 2012; Bigorre et al., 2013), and equated to the 2 and 10 meter values available within reanalyses (T_{2m}, RH_{2m}, WS_{10m}). Both a surface-skin SST (*SST_skin*), to which the surface fluxes are responsive, and a sub-surface foundation SST (*SST_foundation*) are measured.

The ACTIVATE campaign sampled on both sides of the northern Gulf Stream SST gradient (Fig. 1). During February-March, the Gulf Stream meandered to the north, more

noticeable west of 73°W, altering the local SST by more than 4K in places (Fig. 1). The more quiescent synoptic conditions in August-September over a more uniform area support an assessment of boundary-layer depictions less influenced by strong air-sea interactions. Dropsondes were launched from the UC-12 King Air flying at approximately 9 km. Of the 13 King Air flight days in February-March, 2020, 8 coincided with visually-identified CAOs, encompassing 43 of the 59 winter dropsondes (Table S1 lists the individual flight days and their designation as CAO/non-CAO days). Dropsonde circles provide intensive sampling of two CAOs and are the subject of detailed simulations (Li et al., 2021). The 18 August-September flight days include 3 (weaker) CAOs and deployed 107 dropsondes total**.** Neither the buoy nor the dropsonde data are assimilated in either reanalysis or flux product.

The satellite-derived Group for High-Resolution SST (GHRSST), at 9 km spatial resolution, provides spatial context and input to bulk flux calculations. The provided *SST_foundation* at 1m depth is derived from the measured *SST_skin* using a diurnal model.

The ERA5 Reanalysis is arguably the publicly-available reanalysis with the most sophisticated depiction of the cloudy boundary layer. ERA5 has a horizontal grid spacing of 31 km, 137 vertical levels, and an hourly temporal resolution. A systematic bias in the partitioning of ERA5's global wind kinetic energy, with an excessive mean zonal flow coupled with weak meridional flow, is attributed to difficulty in representing high-frequency transient atmospheric events (Rivas and Stoffelen, 2019). The mid-latitude boundary layer maintains a cold and dry bias (Hersbach et al., 2020). The assimilated SST product prior to 2007 was based on the Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST2; Titchner & Rayner, 2014), a 0.25°x0.25° pentad product too coarse to resolve Gulf Stream SST gradients well (Chelton and Risien, 2016). After 2007 it was based on the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012). OSTIA is produced daily on a higher-resolution 0.05^{o}x0.05^{o} grid, and includes satellite microwave measurements, which are less sensitive to the presence of cloud. Systematic differences between the ERA5 SST still remain from other SST climatologies, attributed to spatial resolution (Hersbach et al., 2020). ERA5 provides both *SST_skin* and *SST_foundation*. The 2020 ERA5 *SST_skin* is typically cooler than the *SST_foundation* by 0.2-0.4K (Fig. S2) because of infrared cooling (Fairall et al., 1996; Minnett et al., 2019), with little diurnal warming apparent during the windier, overcast time periods primarily represented here.

MERRA2 spatial resolution is 0.625° x 0.5° longitude by latitude, with 72 vertical levels, and a three-hour temporal resolution, although the surface latent and sensible heat fluxes and lowest-model-level meteorological variables are available at a higher one-hour interval. Previous comparisons to soundings over the Beaufort Sea indicate a MERRA2 warm bias near the surface, and no humidity bias (Rozenhaimer et al., 2018). MERRA2 winds appear too weak, by up to 2 m s^{-1}, in the ACTIVATE region within Molod et al., (2015). MERRA2 only provides *SST_skin*. OAFLUX, at a 1°x1° spatial resolution, provides further insight into the impact of spatial resolution.

Surface fluxes are calculated from the individual dropsondes using the COARE v3.5 parameterization. The COARE v3.0 bulk parameterization values, applied to the buoy T_{2m}, q_{2m} and $WS_{10m,}$ compare well against the CLIMODE buoy direct covariance buoyancy fluxes (Edson et al., 2013; Bigorre et al., 2013), with modifications at wind speeds > 12 m s^{-1} based on the CLIMODE measurements leading to the COARE v3.5 bulk flux algorithm (Edson et al., 2013). Overall this indicates that if the near-surface parameters are known, then the buoyancy fluxes can be estimated with little bias, even within CAOs. Dropsonde inter- or extrapolation provides T_{2m} and q_{2m} and an *SST_skin* is estimated as GHRSST *SST_foundation* + (ERA5 *SST_skin* – ERA5 *SST_foundation*).

The reanalysis comparisons to the dropsondes are nearest in time and space. Instantaneous values captured by the dropsondes from up- and downdrafts will increase the variability of the dropsonde-calculated fluxes beyond those of the coarser-resolution reanalysis fluxes. The comparisons to the CLIMODE buoy data are of the daily- and monthly-mean values, ignoring diurnal variations in wind speed (Dai and Deser, 1997) and near-surface humidity (Clayson and Edson, 2019). Further details can be found in the Supplement.

4. Surface flux representations

a. 2006 CLIMODE

The buoy was located within an SST gradient of approximately 8K over a mere 100 km in January-April of 2006 (Fig. S1). This constitutes a challenging regime for any reanalysis. The monthly-mean SSTs exceed buoy values by up to 5K during February-April, indicating reanalyses depictions of the Gulf Stream that are broader than in nature (Fig. 2 and Fig. S3),

more noticeable as spatial-resolution degrades. Consistent with this, the reanalyses T_{2m} are too warm, and the saturated specific humidity (q_s) too high. The winds are too strong, which physically can be related to the too-warm ocean surface (Small et al., 2008). Monthly-mean reanalyses and OAFLUX buoyancy fluxes differ significantly from the CLIMODE direct covariance values (Fig. S3), most notably in February-March, when buoyancy fluxes exceed the buoy values by 40 to 60 W m^{-2} (see also Table S2), an overestimate of >80%. Since the bulk flux calculations are validated (Edson et al., 2013), the root of the reanalyses flux biases must be their SST representation. This conclusion is in line with Jin and Yu (2013), extended here to newer reanalyses possessing a higher spatial resolution.

Daily-mean differences between the reanalyses/product and CLIMODE buoy values (Fig. 2, Fig. S4 and Table S2) clarify the atmospheric consequences of the SST misrepresentations. The overestimated SST skews reanalysis wind speeds to positive values (Fig. 2f). The ERA5 *SST_foundation* and T_{2m} deviate the least from the buoy values (Fig. 2a, b), even though the MERRA2 *SST_skin* values should in theory be cooler (Fig. S2). Both reanalyses match the buoy wind speeds well (Fig. 2f, Table S2). The surface q_s is elevated for all reanalyses/product, as expected. A dry bias in ERA5's q_{2m} contrasts with a moist bias for MERRA2's q_{2m}, both by about 1 g kg^{-1}. In combination, the air-sea thermodynamic differences are smaller for MERRA2 on most days, compared to ERA5 (Fig. 2g and h, Table S2), compensating for MERRA2's poorer T_{2m} (Fig. 2a). This allows the MERRA2 buoyancy fluxes to ultimately compare better to the CLIMODE values, than the ERA5 values (Fig. 2e). Overall, this comparison suggests ERA5 provides a more accurate depiction of the Gulf Stream near-surface meteorology, likely in part because of an improved resolution, but compensations within MERRA2 model physics may be improving the fluxes, if for the wrong reasons. OAFLUX, with the coarsest resolution, has flux, SST and wind speed values that diverge the most of the three products from CLIMODE buoy values (Fig. S3; Fig. 2b, e, f; Table S2).

During the summer months, the SST differences are smaller and more evenly distributed about zero (Fig. 2j and S4). Buoyancy fluxes remain consistently overestimated (Fig. 2m), most noticeable by ERA5 because of its too-cool T_{2m} and too-dry (by 1-2 g/kg) q_{2m}. The too-dry ERA5 q_{2m} bias for both seasons indicates a common bias source. In contrast, a too-warm T_{2m} in winter and too-cool T_{2m} in summer suggests differing underlying causes.

b. 2020 ACTIVATE

ERA5 *SST_foundation* exceed GHRSST values by up to 2K during February-March of 2020 at the Gulf Stream boundaries, and a slight underestimation of the cooler southward-flowing coastal Labrador Current temperatures is also evident (Fig. 1d). The ERA5 SST bias exists even though the assimilated SST product is of a similar (slightly finer) spatial resolution as GHRSST (0.05° versus 9 km), reflecting the coarsening needed to match the ERA5 resolution of 31 km (Hersbach et al., 2020).

The hourly-mean ERA5 reanalysis sensible and latent heat fluxes overestimate during the more severe CAOs, by up to 100 W m^{-2} (Fig. 3a) or more for the latent heat fluxes (Fig. 3b). The cause is most clearly linked to wind speed overestimates. In contrast, the MERRA2 fluxes always underestimate, because of underestimates in the air-sea temperature (Fig. 3c) and humidity (Fig. 3d) differences, and weaker MERRA2 wind speeds (Fig. 3e). Differences in temporal/spatial resolution (instantaneous versus hourly-mean values over a larger spatial domain) seem unlikely to explain the MERRA2 underestimates, given that these are systematic biases, and instead point to a near-surface boundary layer that is too close in thermodynamic equilibrium with the ocean.

In August-September 2020, both reanalyses slightly underestimate the fluxes relative to those calculated from the dropsondes, with ERA5 performing better than MERRA2 (Table S3). Air-sea humidity differences are more realistically captured by ERA5 than by MERRA2. MERRA2, similar to the winter months, consistently underestimates all inputs into the flux calculations, although the biases are small (Table S3).

5. Thermodynamic Vertical Structure

Figure 4 compares the ERA5 and MERRA2 vertical atmospheric structure to the dropsonde-derived mean potential temperature (θ), *RH* and *q*, and wind speed (total, zonal and meridional) for February-March and August-September of 2020, while Fig. S5 indicates the differences more explicitly. The reanalyses capture the main features of the lower tropospheric structure. Consistent with the near-surface analysis, the wintertime ERA5 boundary layer is slightly too cold and too dry, with *RH* and *q* averages indicating underestimates of 5% (ranging up to 30%) and 0.5 g kg^{-1} (ranging up to 2 g kg^{-1}), respectively. Locations with ERA5-SST – GHRSST > 1K reveal mean ERA5 boundary layer θ profiles that are 0.2K warmer, ranging up to

1 to 2K. Where ERA5-SST – GHRSST < -1K, the ERA5 θ profiles are almost 1K cooler than the dropsonde values (inset plot in Fig. 4a).

The mean ERA5 wind biases are small, with a slight overestimation (1 m s^{-1}, or 10%) that primarily comes from the zonal component. The lower free troposphere in ERA5 does not fully resolve the observed structure (e.g., the elevated moisture layer between 800-750 hPa), but the main inversion top at approximately 850 hPa, identified using an *RH* threshold, is adequately captured. The winter MERRA2 thermodynamic structure compares more closely to the *in-situ* values (seen more clearly in Fig. S5). Interestingly, the sign of the MERRA2 wind bias contrasts with that from ERA5. An underestimate of the near-surface zonal winds increases with altitude, suggesting a downward momentum transport may explain the near-surface bias.

The ERA5 wind biases are smaller in August-September, while ERA5 θ remains depressed by 0.2 - 0.3 K (ranging up to 3K) and q also remains biased low, by up to 0.5 g kg^{-1}. These compensate to generate realistic *RH* values near the surface (Fig. S5). Specific humidity underestimates above the surface-based mixed layer are larger for MERRA2 than ERA5, permeating into the relative humidity. Both ERA5 and MERRA2 struggle with capturing the cloud layer between 900-800 hPa (Fig. 4h inset). This is not linked to a pronounced bias in the winds for ERA5, while MERRA2 winds are clearly too weak above 1 km.

In contrast to ERA5, the *RH* and q-mean MERRA2 profiles agree well with the dropsondes during February-March, while the zonal winds are consistently weaker, by 1-2 m s^{-1} near the surface, increasing (mostly) with altitude. During the late summer, MERRA2 is more likely to be drier within 0.4-2.0 km than the *in-situ* measurements, indicating the critical relative humidity threshold for cloud production may be set too low then (Molod et al., 2015).

During February-March, both ERA5 and MERRA2 capture the inversion height of approximately 1.7 km reasonably well (estimated from the *RH* profiles). During August-September, the inversion is naturally lower, at approximately 1.1 km (similarly estimated). Both reanalyses often fail to capture the cloud layer in late summer. The mean MERRA2 boundary layer height is lower than that from ERA5, with a drier cloud layer. The lower boundary layer height is even more pronounced after the few September CAO cases are excluded (not shown), indicating the issue may be a similar difficulty in representing stable boundary layers as for ERA5. The MERRA2 winds, both zonal and meridional, are also weaker.

5 Discussions and Conclusions

Hersbach et al. (2020) note that the too-dry ERA5 boundary layer, evident in all seasons and in both the CLIMODE and ACTIVATE comparisons, coincides with a warming of the lower troposphere and with the advent of microwave imagers. These are shown to warm and dry ERA5 at 850 hPa over the ocean (Geer et al., 2017). The exact mechanisms do not yet appear to be known (Hersbach et al., 2020); the assimilation of microwave radiances, although providing additional information, can nevertheless not fully constrain the thermodynamic profile (e.g., Pincus et al., 2017; Zhang et al., 2018), introducing understandable trade-offs.

The too-strong wintertime westerlies generate an anomalous wind convergence near the surface (Rivas and Stoffelen, 2019), which should act to raise the boundary layer height, all else being equal. The reasonable depiction of the ERA5 wintertime boundary layer depth, whereas the late summer cloudy boundary layer is too shallow, may reflect a conscious choice at numerical weather prediction centers to artificially enhance the turbulent diffusion in stable conditions, towards improving the depiction of synoptic cyclones (Sandu et al., 2013). The ACTIVATE campaign is selectively sampling CAO conditions during its winter campaigns, for which the ERA5 turbulent diffusion choices are optimized. The late summer time period, when the north Atlantic sea level pressure high extends further west (Painemal et al., 2021), provides conditions in which both reanalyses have more difficulty in maintaining a cloudy stable layer. The artificial enhancement in the ERA5 diffusion parameters was also intended to improve a near-surface cold temperature bias (by encouraging the entrainment of warmer air aloft); we find a small (~0.2K) ERA5 cold temperature bias still remains during both winter and late summer.

The length scale of ocean mesoscale eddies is 20-30 km at the latitude of the Gulf Stream, set to first-order by the Rossby radius. ERA5 possesses the horizontal grid spacing best able to represent the majority of the ocean mesoscale activity of the three products examined, though still missing the smallest eddies. A wintertime western boundary current that is too wide in the reanalyses could imply that the boundary layer adjustment for air coming from the west might be affected earlier within the reanalysis than in nature. CAO air flows first over cooler coastal waters north of 35°N generated by the Labrador Current, whose ERA5 reanalysis temperatures are too cool, potentially further energizing the adjustment process of the boundary layer to the warmer Gulf Stream waters. In addition, all of the biases in ERA5 contribute to

exaggerating the surface heat fluxes during CAOs, which will also contribute to elevating the inversion. Perhaps because of these characteristics, the wintertime ERA5 boundary layer depth, thermodynamic and dynamic structure is broadly representative of the observations. ERA5 vertical motion fields have also been shown to compare well to those derived from the two dropsonde circles (Li et al., 2021), lending further confidence in the ability of ERA5 to depict the strongly-forced cold-air outbreak regime. A correct boundary layer depth is a critical parameter for shallow clouds, as the depth affects the coupling to the ocean surface. The robustness of the bias in ERA5 q suggests an observationally-determined correction factor could be applied, improved as more dropsonde data become available. A similar approach could be adopted for the too-cool ERA5 temperature bias. For those LES studies assuming a Lagrangian perspective, additional ACTIVATE dropsonde data will also support analysis of how much reanalysis profiles deviate from observations as a function of distance from shore. Interestingly, the biases in the MERRA2 reanalysis are different from those in ERA5, tending to too-weak surface fluxes. Nevertheless the boundary layer depth depiction is similar: approximately realistic during the winter, and too shallow during the summer. The momentum transport to the surface by ERA5 can increase the surface wind speed, deepening the boundary layer more quickly and encouraging a faster cloud transition, than in nature (Saggiaroto et al., 2020). Future work will incorporate space-based lidar and radar data and *in-situ* measurements to evaluate the CAO cloud structure evolution over the Gulf Stream.

Acknowledgments and Data Availability

This work was supported by NASA grant 80NSSC19K0390 in support of ACTIVATE, a NASA Earth Venture Suborbital-3 (EVS-3) investigation funded by NASA's Earth Science Division and managed through the Earth System Science Pathfinder Program Office. The Pacific Northwest National Laboratory (PNNL) is operated for the US Department of Energy (DOE) by Battelle Memorial Institute under Contract DE-AC06-76RLO 1830. Microwave OI SST data are produced by Remote Sensing Systems, available at www.remss.com. ERA5 data are available through https://cds.climate.copernicus.eu/. MERRA-2 data are available at https://disc.gsfc.nasa.gov/. CLIMODE buoy measurements are available at http://www.opal.sr.unh.edu/data/airsea_flux.shtml. The OAFlux data are available through http://oaflux.whoi.edu/.

References

Abel, S. J., Boutle, I. A., Waite, K., Fox, S., Brown, P. R. A., Cotton, R., Lloyd, G., Choularton, T. W., & Bower, K. N. (2017). The Role of Precipitation in Controlling the Transition from

Stratocumulus to Cumulus Clouds in a Northern Hemisphere Cold-Air Outbreak, *Journal of the Atmospheric Sciences*, *74*(7), 2293-2314.

Bane, J. M., Jr., & Osgood, K. E. (1989). Wintertime air-sea interaction processes across the Gulf Stream. *Journal of Geophysical Research*, 94 (C8), 10 755–10 722. doi: 10.1029/JC094iC08p10755.

Bigorre, S., Weller, R. A., Lord, J., Edson, J. B., & Ware, J. D. (2013). A surface mooring for air–sea interaction research in the Gulf Stream. Part 2: Analysis of the observations and their accuracies. *Journal of Atmosheric and Oceanic Technology*, 30, 450–469, doi:10.1175/ JTECH-D-12-00078.1.

Bony, S., & Stevens, B. (2019). Measuring Area-Averaged Vertical Motions with Dropsondes, *Journal of the Atmospheric Sciences*, *76*(3), 767-783

Bosilovich, M. G. and Coauthors. (2015). *MERRA-2: Initial evaluation of the climate. Technical Report Series on Global Modeling and Data Assimilation*, Vol. 43, NASA Tech. Rep. NASA/TM– 2015–104606, 139 pp. [Available online at https://gmao.gsfc. nasa.gov/pubs/docs/Bosilovich803.pdf.]

Chelton, D. B., & Risien, C.M. (2016). *Zonal and Meridional Discontinuities and Other Issues with the HadISST1.1 Dataset*. Oregon State University, Corvallis, Oregon, Technical Report. doi: 10.13140/RG.2.1.4503.0168.

Clayson, C. A., & Edson, J. B. (2019). Diurnal surface flux variability over western boundary currents. *Geophysical Research Letters*, 46, 9174–9182. Doi: 10.1029/2019GL082826

Dai, A., & Deser, C. (1997). Diurnal and semidiurnal variations in global surface wind and divergence fields. *Journal of Geophysical Research*, 104, 31,109–31,125

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sensing of Environment*, 116, 140–158, doi:10.1016/j.rse.2010.10.017.

Edson, J. B., Jampana, V., Weller, R.A., Bigorre, S.P., Plueddemann, A.J., Fairall, C.W., Miller, S.D., Mahrt, L., Vickers, D., & Hersbach, H. (2013). On the exchange of momentum over the open ocean. *Journal of Physical Oceanography*, 43, 1589–1610. doi:10.1175/ JPO-D-12-0173.1.

Fairall, C.W., Bradley, E.F., Godfrey, J.S. , Edson, J.B., Young, G.S. and Wick, G.A. (1996) Cool skin and warm layer effects on the sea surface temperature. *J. Geophys. Res., 101,* 1295-1308.

Fairall C. W., Bradley, E., Hare, J., Grachev, A., & Edson, J. (2003). Bulk parameterization of air–sea fluxes: updates and verification for the COARE algorithm. *Journal of Climatology*, 16, 571–591. doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2

Field, P. R., Cotton, R. J., McBeath, K., Lock, A. P., Webster, S., & Allan, R. P.(2014). Improving a convection-permitting model simulation of a cold air outbreak. *Quarterly Journal of Royal Meteorological Society*, 140, 124–138, https://doi.org/10.1002/qj.2116.

Field, P. R., & Heymsfield, A. J. (2015). Importance of snow to global precipitation. *Geophysical Research Letter*, 42, 9512–9520, doi:10.1002/2015GL065497.

Fletcher, J. K., Mason, S. L., & Jakob, C. (2016a). The climatology, meteorology, and boundary layer structure of marine cold air outbreaks in both hemispheres. *Journal of Climate*, 29, 1999–2014, https://doi.org/10.1175/JCLI-D-15-0268.1.

Fletcher, J. K., Mason, S. L., & Jakob, C. (2016b). A climatology of clouds marine cold air outbreaks in both hemispheres. *Journal of Climate*, 29, 6677–6692, https://doi.org/10.1175/JCLI-D-15-0783.1.

Geer, A.J., Baordo, F., Bormann, N., English, S.J., Kazumori, M., Lawrence, H., Lean, P., Lonitz, K. & Lupu, C. (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quarterly Journal of the Royal Meteorological Society*, 143, 3189–3206. doi:10.1002/qj.3172

Grossman, R. L.& Betts, A. K. (1990). Air–sea interaction during an extreme cold air outbreak from the eastern coast of the United States. *Monthly Weather Review*, 118, 324–342.

Hersbach, H., Bell, B., Berrisford, P., et al. (2020). The ERA5 global analysis. *Quarterly Journal of Royal Meteorological Society*, 146, 1999-2049. doi:10.1002/qj.3803

Honnert, R., Efstathiou, G., Beare, R., Ito, J., Lock, A., Neggers, R., et al. (2020). The atmospheric boundary layer and the "gray zone" of turbulence: A critical review.*Journal of Geophysical Research: Atmospheres*, 125, e2019JD030317. doi.org/10.1029/2019JD030317

Jin, X., and L. Yu (2013), Assessing high-resolution analysis of surface heat fluxes in the Gulf Stream region, *J. Geophys. Res. Oceans*, 118, 5353–5375, doi:10.1002/jgrc.20386

Kolstad, E.W., Bracegirdle, T. J., & Seierstad, I. A. (2009). Marine cold-air outbreaks in the North Atlantic: Temporal distribution and associations with large-scale atmospheric circulation. *Climate Dynamics*, 33, 187–197, https://doi.org/10.1007/s00382-008-0431-5.

Li, X.-Y., Wang, H., Chen, J., Endo, S., George, G., Cairns, B., et al., (2021). Large-eddy simulations of marine boundary-layer clouds associated with cold air outbreak during the ACTIVATE campaign – Part 1: Case setup and sensitivities to large-scale forcings. *Journal of Atmospheric Science*. Submitted.

Liu, J., Xie, S., Norris, J. R., & Zhang, S. (2014). Low-Level Cloud Response to the Gulf Stream Front in Winter Using CALIPSO. *Journal of Climate*, *27*(12), 4421-4432.

Marshall, J., Andersson, A., Bates, N., Dewar,W., Doney, S., Edson, J., et al., (2009). The CLIMODE field campaign: Observing the cycle of convection and restratification over the Gulf Stream. *Bull. Amer. Meteor. Soc., 90*, 1337-1350.

Masunaga, R., Nakamura, H., Miyasaka, T., Nishii, K., & Tanimoto, Y. (2015). Separation of Climatological Imprints of the Kuroshio Extension and Oyashio Fronts on the Wintertime Atmospheric Boundary Layer: Their Sensitivity to SST Resolution Prescribed for Atmospheric Reanalysis. *Journal of Climate*, 28, 1764–1787. doi:10.1175/JCLI-D-14-00314.1.

McCoy, I., Wood, R., & Fletcher, J. K. (2017). Identifying meteorological controls on open and closed mesoscale cellular associated with marine cold air outbreaks. *Journal of Geophysical Research,* 122, 11,678–11,702, https://doi.org/10.1002/2017JD027031, 2017

Minnett, P. J., A. Alvera-Azcárate, T.M. Chin, G.K. Corlett, C.L. Gentemann, I. Karagali, et al., (2019). Half a century of satellite remote sensing of sea-surface temperature, *Remote Sensing of Environment*, 233, doi: 10.1016/j.rse.2019.111366.

Minobe, S., Kuwano-Yoshida, A., Komori, N., Xie, S.P., & Small, R.J. (2008). Influence of the Gulf Stream on the troposphere. *Nature*, 452, 206–209, doi: 10.1038/nature06690.

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. *Geoscientific Model Development,* 8, 1339–1356, https://doi.org/10.5194/gmd-8-1339-2015, 2015.

Moore, G. W. K., & Renfrew, I.A. (2002). An assessment of the surface turbulent heat fluxes from the NCEP–NCAR reanalysis over the western boundary currents. *Journal of Climate*, 15, 2020–2037. doi: 10.1175/1520-0442(2002)015<2020:AAOTST>2.0.CO;2

Mülmenstadt, J., Sourdeval, O., Delano., J., & Quaas. J. (2015). Frequency of occurrence of rain from liquid-, mixed-, and ice-phase clouds derived from A-Train satellite retrievals. *Geophysical Research Lett*er 42, 6502–6509, doi:10.1002/2015GL064604.

Pincus, R., Beljaars, A., Buehler, S.A., Kirchengast, G., Ladstaedter, F., Whitaker, J. S. (2017). The Representation of Tropospheric Water Vapor Over Low-Latitude Oceans in (Re-)analysis: Errors, Impacts, and the Ability to Exploit Current and Prospective Observations. *Surv Geophys* 38**,** 1399–1423. doi:10.1007/s10712-017-9437-z

Plagge, A., Edson, J. B., & Vandemark, D. (2016). In Situ and Satellite Evaluation of Air–Sea Flux Variation near Ocean Temperature Gradients. *Journal of Climate*, *29*(4), 1583-1602.

Painemal, D., Corral, A. F., Sorooshian, A., Brunke, M. A., Seethala, C.,, Gorooh, V. A. et a., (2021). An Overview of Atmospheric Features Over the Western North Atlantic Ocean and North American East Coast – Part 2: Circulation, Boundary Layer, and Clouds. *Journal of Geophysical Research,* 126, e2020JD033423. doi:10.1029/2020JD033423

Remote Sensing Systems (REMSS) (2008). *GHRSST Level 4 MW_IR_OI Global Foundation Sea Surface Temperature analysis version 5.0 from REMSS (GDS versions 1 and 2)*. NOAA

National Centers for Environmental Information. doi: gov.noaa.nodc:GHRSST-MW_IR_OI-REMSS-L4-GLOB

Reynolds, R.W. & Smith, T.M. (1994). Improved Global Sea Surface Temperature Analysis using Optimum Interpolation. *Journal of Climate*, 7, 929-948. doi: 10.1175/1520-0442(1994)007

Rivas, M. B., & Stoffelen, A. (2019). Characterizing ERA-Interim and ERA5 surface wind biases using ASCAT. *Ocean Science*, 15, 831-851. doi: 10.5194/os-15-831-2019

Rozenhaimer, M., Barton, N., Redemann, J., Schmidt, S., LeBlanc, S., Anderson, B., et al. (2018). Bias and sensitivity of boundary layer clouds and surface radiative fluxes in MERRA-2 and airborne observations over the Beaufort Sea during the ARISE campaign. *Journal of Geophysical Research,* 123, 6565–6580. Doi:10.1029/2018JD028349

Saggiorato, B., Nuijens, L., Siebesma, A. P., de Roode, S., Sandu, I.,& Papritz, L. (2020). The influence of convective momentum transport and vertical wind shear on the evolution of a cold air outbreak. *Journal of Advances in Modeling Earth System*, 12, https://doi.org/10.1029/2019MS001991.

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58, e2019RG000678. https://doi.org/10.1029/2019RG000678.

Skyllingstad, E. D., & Edson, J. B. (2009). Large-eddy simulation of moist convection during a cold-air outbreak over the Gulf Stream. *Journal of the Atmospheric Sciences*, 66(5), 1274–1293. https://doi.org/10.1175/2008JAS2755

Small, R. J., & Coauthors. (2008). Air–sea interaction over ocean fronts and eddies. *Dynamics of Atmosphere and Oceans*, 45, 274–319, doi:10.1016/j.dynatmoce.2008.01.001.

Sorooshian, A., Anderson, B., Bauer, S. E., Braun, R. A., Cairns, B., Crosbie, E., et al., (2019). Aerosol–cloud–meteorology interaction airborne field investigations: Using lessons learned from the U.S. West Coast in the design of ACTIVATE off the U.S. East Coast. *Bulletin of the American Meteorological Society*, 100(8), 1511– 1528. https://doi.org/10.1175/BAMS-D-18-0100.1.

Titchner, H. A., & Rayner, N.A. (2014). The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *Journal of Geophysical Research Atmospheres, 119,* 2864-2889. doi: 10.1002/2013JD020316.

Tornow, F., Ackerman, A. S., & Fridlind, A. M. (2021). Preconditioning of overcast-to-broken cloud transitions by riming in marine cold air outbreaks. *Atmospheric Chemistry and Physics Discussions*, https://doi.org/10.5194/acp-2021-82.

Weller, R.A., Bigorre, S.P., Lord, J., Ware, J.D., & Edson, J.B. (2012). A Surface Mooring for Air–Sea Interaction Research in the Gulf Stream. Part I: Mooring Design and Instrumentation.

Journal of Atmosheric and Oceanic Technology, 29 (9): 1363–1376. doi:10.1175/JTECH-D-12-00060.1

Yu, L., Jin, X., & Weller, R.A. (2008). *Multidecade Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables*. Woods Hole Oceanographic Institution, OAFlux Project Technical Report. OA-2008-01, 64pp. Woods Hole. Massachusetts.

Zelinka, M., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S., & Taylor, K. E. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letter*, 47, https://doi.org/10.1029/2019GL085782, e2019GL085782, 2020.

Zhang, J., Zuidema, P., Turner, D., & Cadeddu, M. (2018). Surface-based microwave humidity retrievals over the equatorial Indian Ocean: Applications and challenges. *Journal of Applied Meteorology and Climatology*, 57, 1765-1782, doi:10.1175/JAMC-D-17-0301.1

Figures

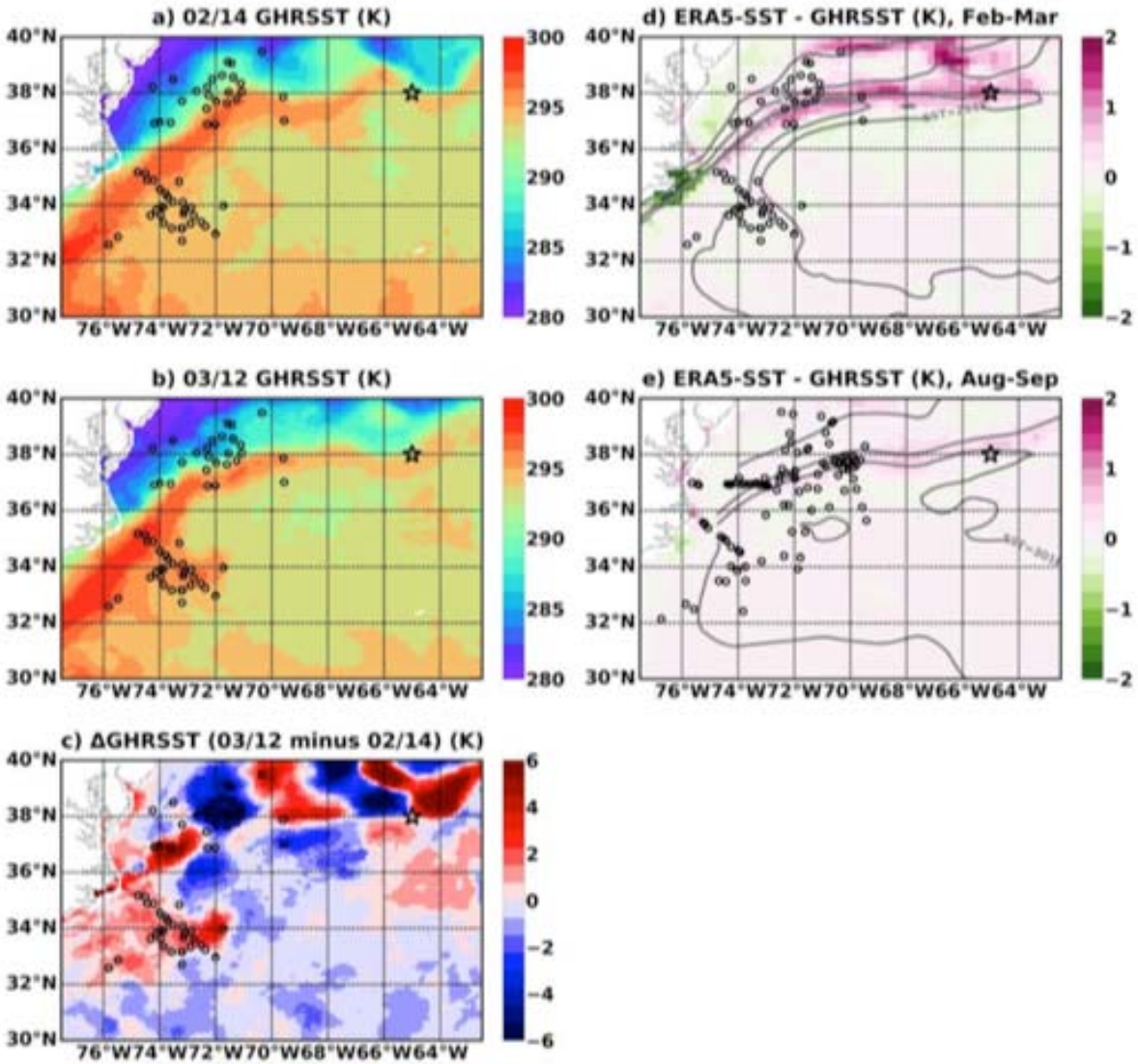


Figure 1. Sea surface temperatures (GHRSST) on a) 14 February 2020, b) 12 March 2020, and c) 12 March - 14 February, 2020. Open circles indicate ACTIVATE dropsonde locations, and the star denotes the CLIMODE buoy location. (ERA5-SST – GHRSST) differences for d) February-March 2020 and e) August-September 2020. Gray contours in d) and e) correspond to ERA5 SSTs of 286, 290, 294 and 295K in panel d) and 301K in panel e).

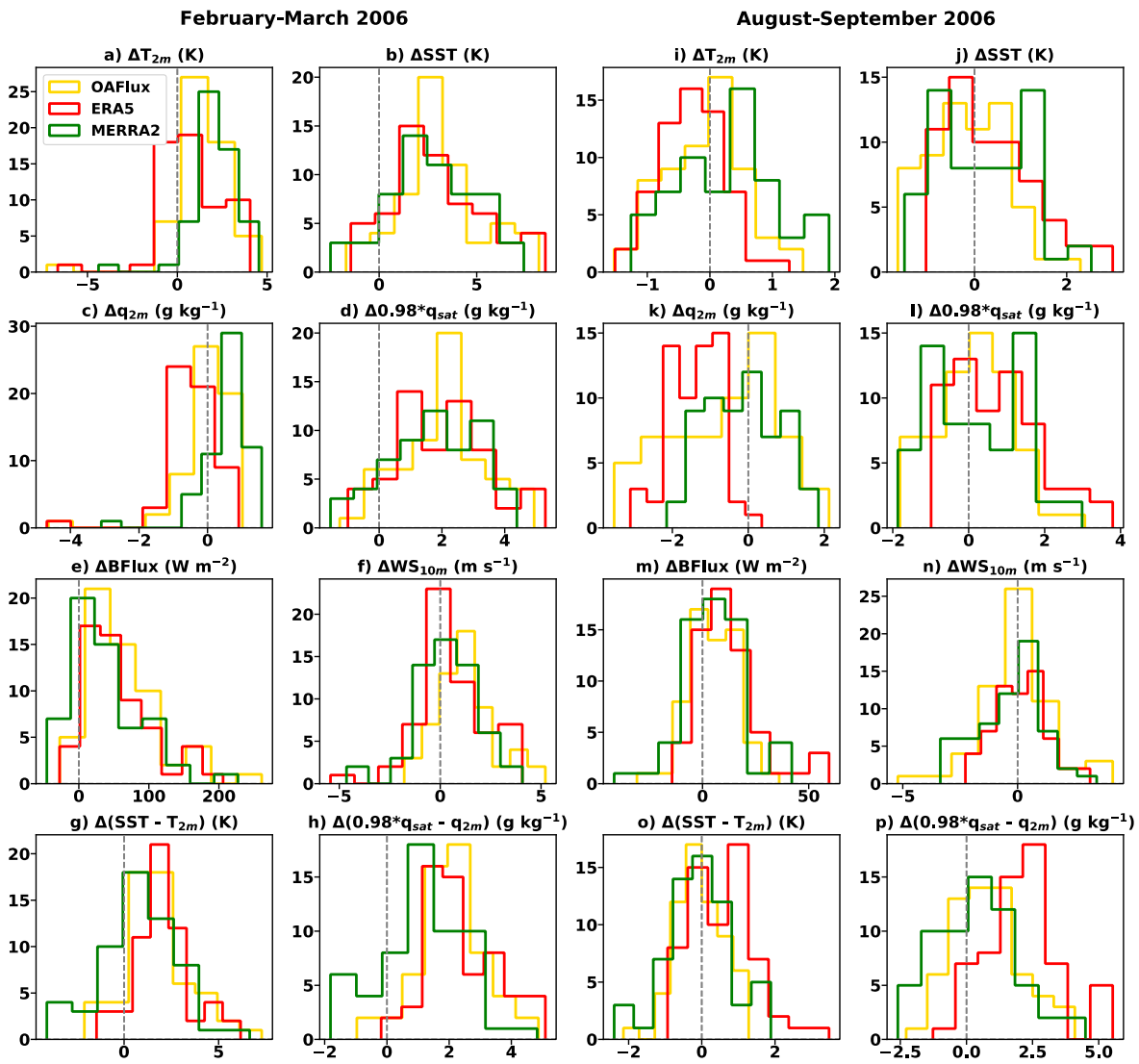


Figure 2. Comparison of ERA5 (red), MERRA2 (green), and OAFLUX (yellow) daily-mean differences from CLIMODE buoy measurements (reanalysis-buoy) for February-March 2006 for a) near-surface air temperature, b) sea surface temperature, c) near-surface specific humidity, d) saturated specific humidity at the SST, e) buoyancy flux, f) wind speed at 10m, and the air-sea g) temperature and h) humidity differences. i)-p): same as a)-h) but for August-September 2006. Note changes in x-scale range. MERRA2 SST is a skin value, while the buoy, ERA5, and OAFLUX SSTs are foundation SST. OAFLUX WS_{10m} is the neutral wind speed.

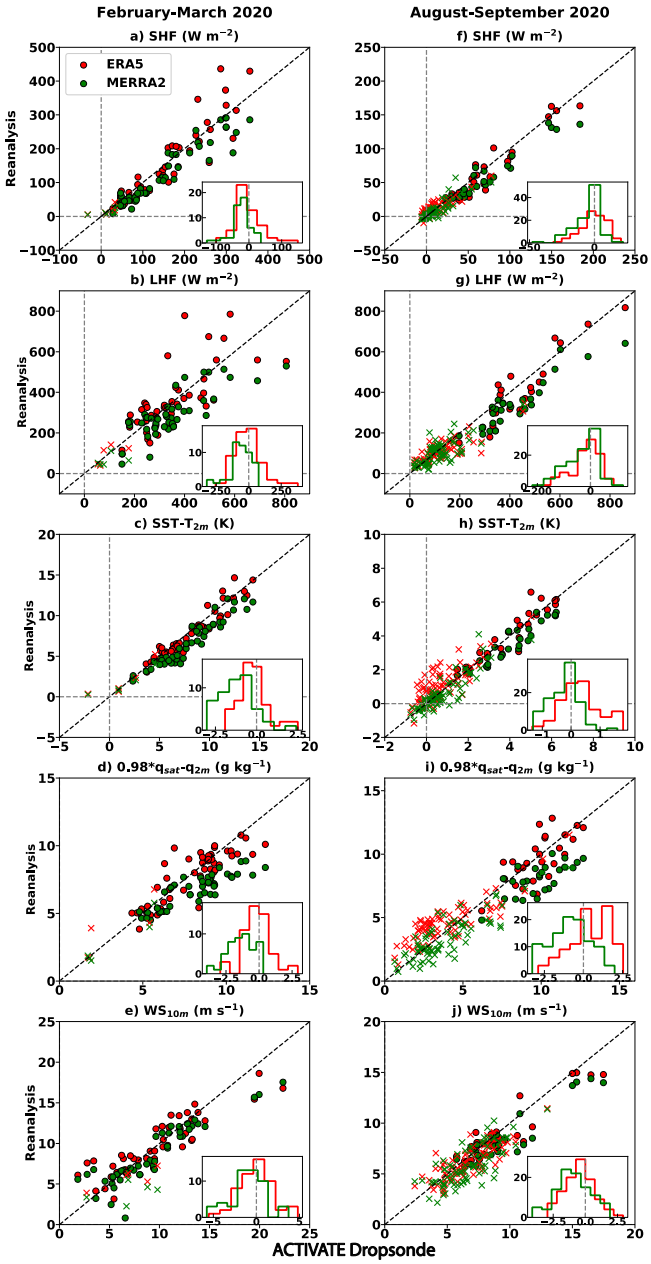


Figure 3. Comparison of nearest-in-space-and-time hourly-mean ERA5 (red) and three-hourly-mean MERRA2 (green) reanalysis to instantaneous ACTIVATE dropsonde-calculated values for February-March 2020 of a) sensible heat flux, b) latent heat flux, calculated using the COARE v3.5 algorithm, c) $SST\text{-}T_{2m}$, d) $0.98*q_s\text{-}q_{2m}$, and e) WS_{10m}. f)-j): same as a)-e) but for August-September 2020; note change in range on both axes. Insets within each panel are histograms of the (reanalysis-dropsonde) differences. Filled circles represent cold-air outbreak conditions, and 'x' markers signify non-CAO conditions, using θ_{SST_skin} - θ_{900hPa} > 0 to define whether an individual dropsonde represented a CAO (see supplement). GHRSST values are corrected to represent a 'skin' SST.

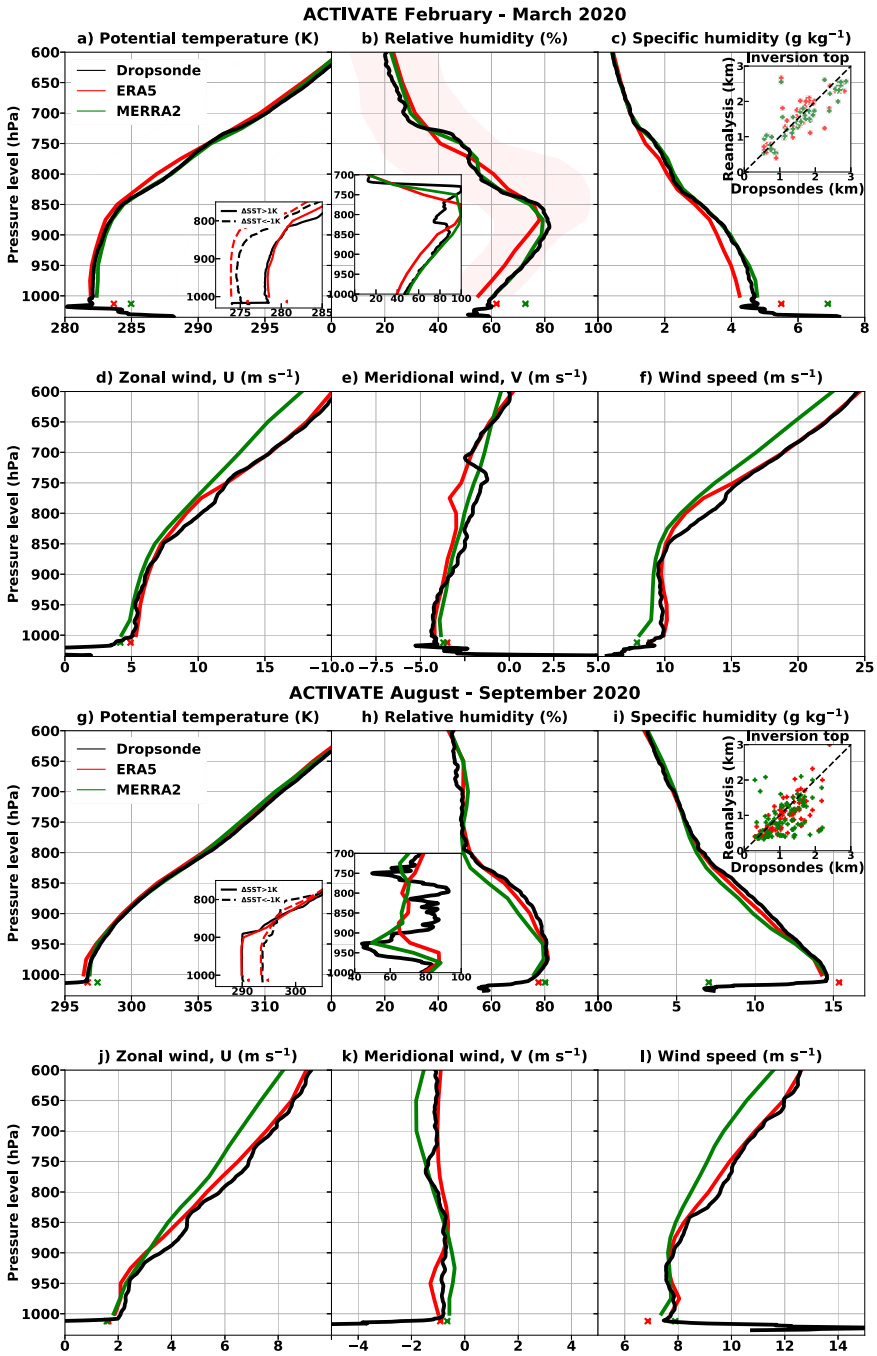


Figure 4. Mean vertical profiles from dropsondes, ERA5, and MERRA2 of a) potential temperature θ, b) relative humidity, c) specific humidity, d) zonal wind (U), e) meridional wind (V), and f) wind speed for February-March 2020. Colored shading indicates the standard deviation. Small 'x' markers indicate θ_{2m}, RH_{2m} and q_{2m}, and U, V and *WS* at 10m. f)-l): same as a)-f) but for August-September 2020. Insets in a) and g) represent the mean θ profiles for ERA5-SST – GHRSST > 1K and ERA5-SST – GHRSST < -1K. The inset profile in panel b) is from 28 February (profile#24) and in panel h) from 20 August (profile#12). Insets in c) and i) indicate inversion top heights estimated from the *RH* profiles.

Figure 1.

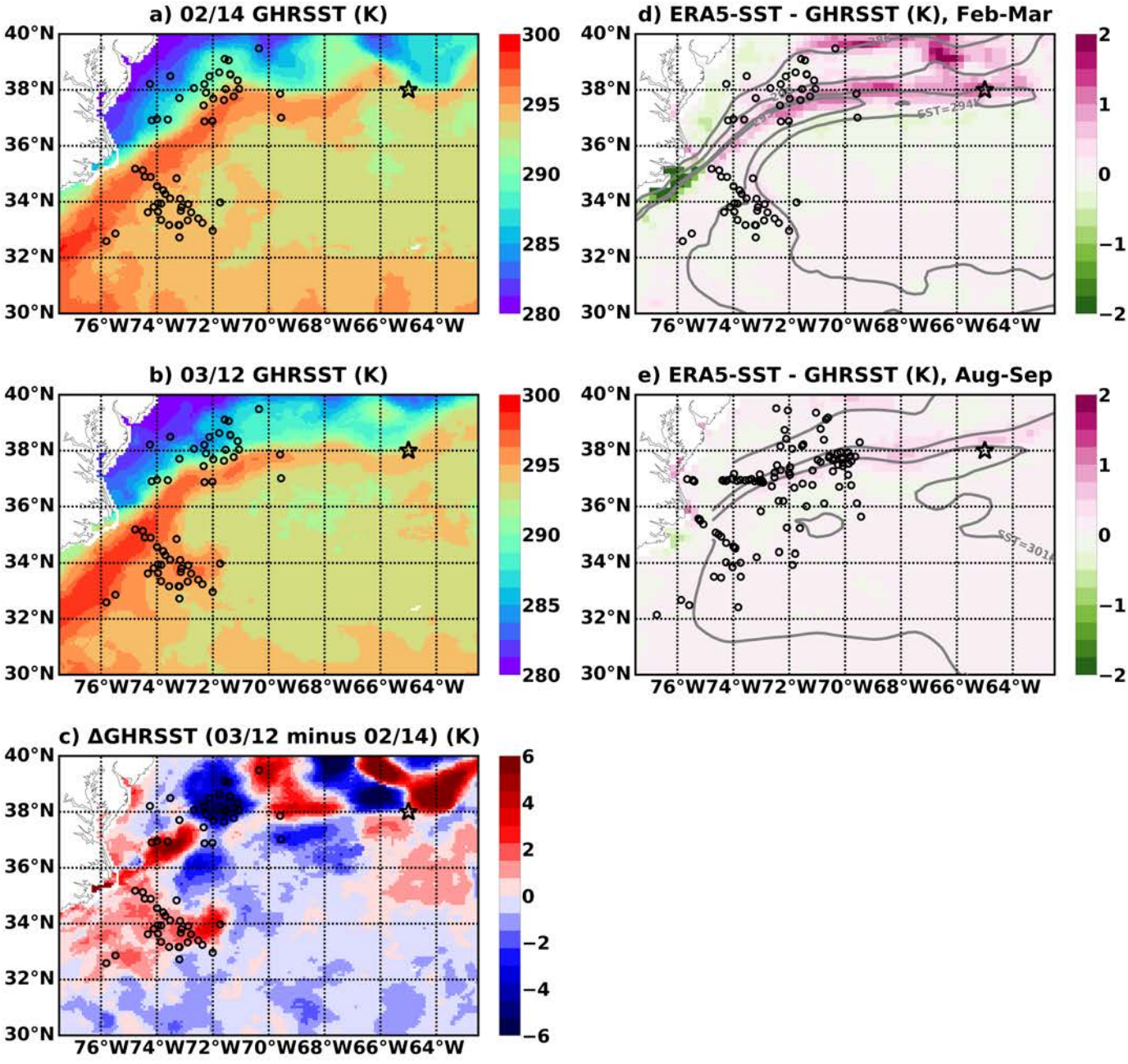

a) 02/14 GHRSST (K)
b) 03/12 GHRSST (K)
c) ΔGHRSST (03/12 minus 02/14) (K)
d) ERA5-SST - GHRSST (K), Feb-Mar
e) ERA5-SST - GHRSST (K), Aug-Sep
SST=294K
SST=301K
40°N
38°N
36°N
34°N
32°N
30°N
76°W74°W72°W70°W68°W66°W64°W
300
295
290
285
280
6
4
2
0
−2
−4
−6
2
1
0
−1
−2

Figure 2.

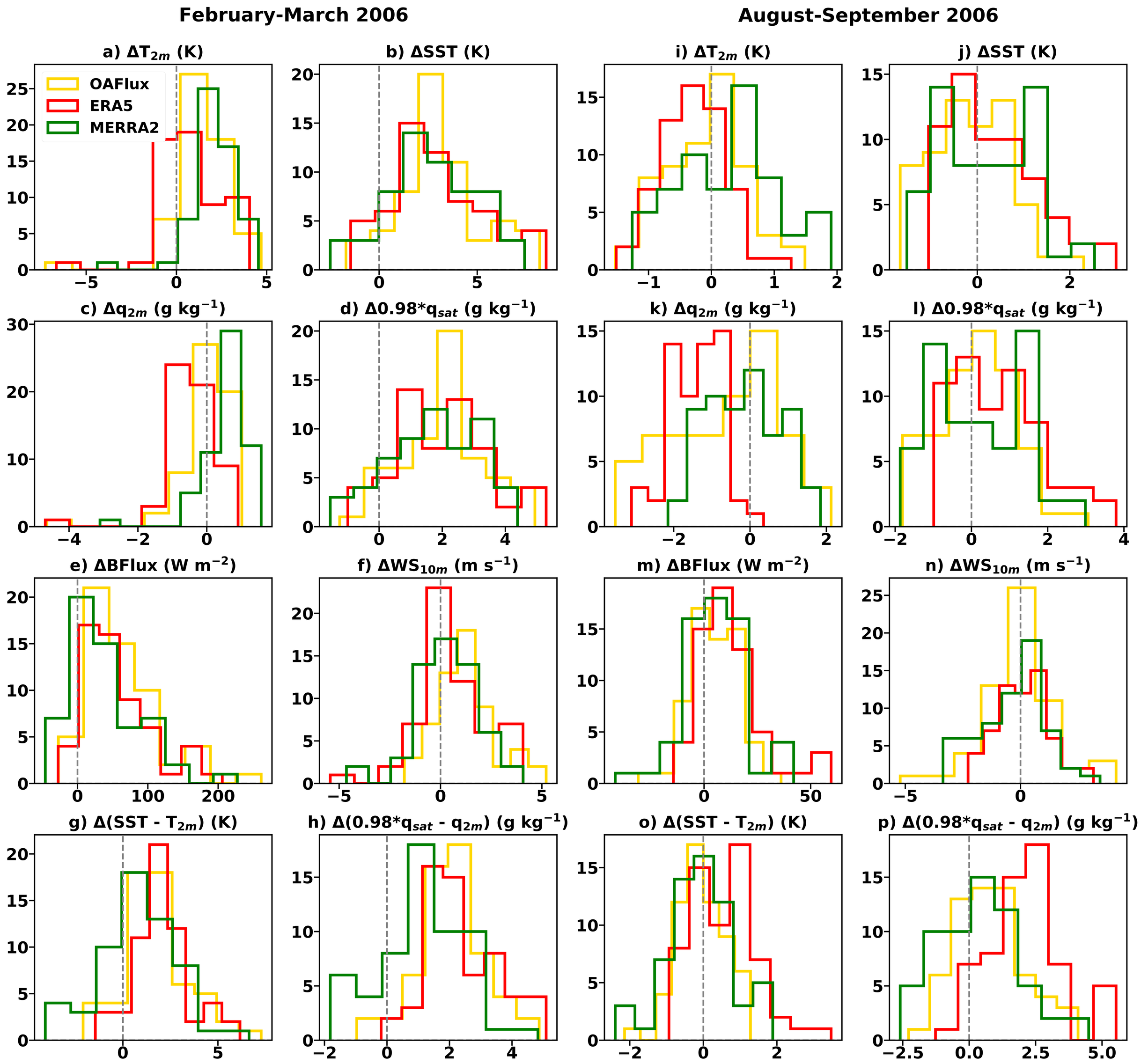

February-March 2006
August-September 2006
a) ΔT2m (K)
b) ΔSST (K)
i) ΔT2m (K)
j) ΔSST (K)
OAFlux
ERA5
MERRA2
c) Δq2m (g kg−1)
d) Δ0.98*qsat (g kg−1)
k) Δq2m (g kg−1)
l) Δ0.98*qsat (g kg−1)
e) ΔBFlux (W m−2)
f) ΔWS10m (m s−1)
m) ΔBFlux (W m−2)
n) ΔWS10m (m s−1)
g) Δ(SST - T2m) (K)
h) Δ(0.98*qsat - q2m) (g kg−1)
o) Δ(SST - T2m) (K)
p) Δ(0.98*qsat - q2m) (g kg−1)

Figure 3.

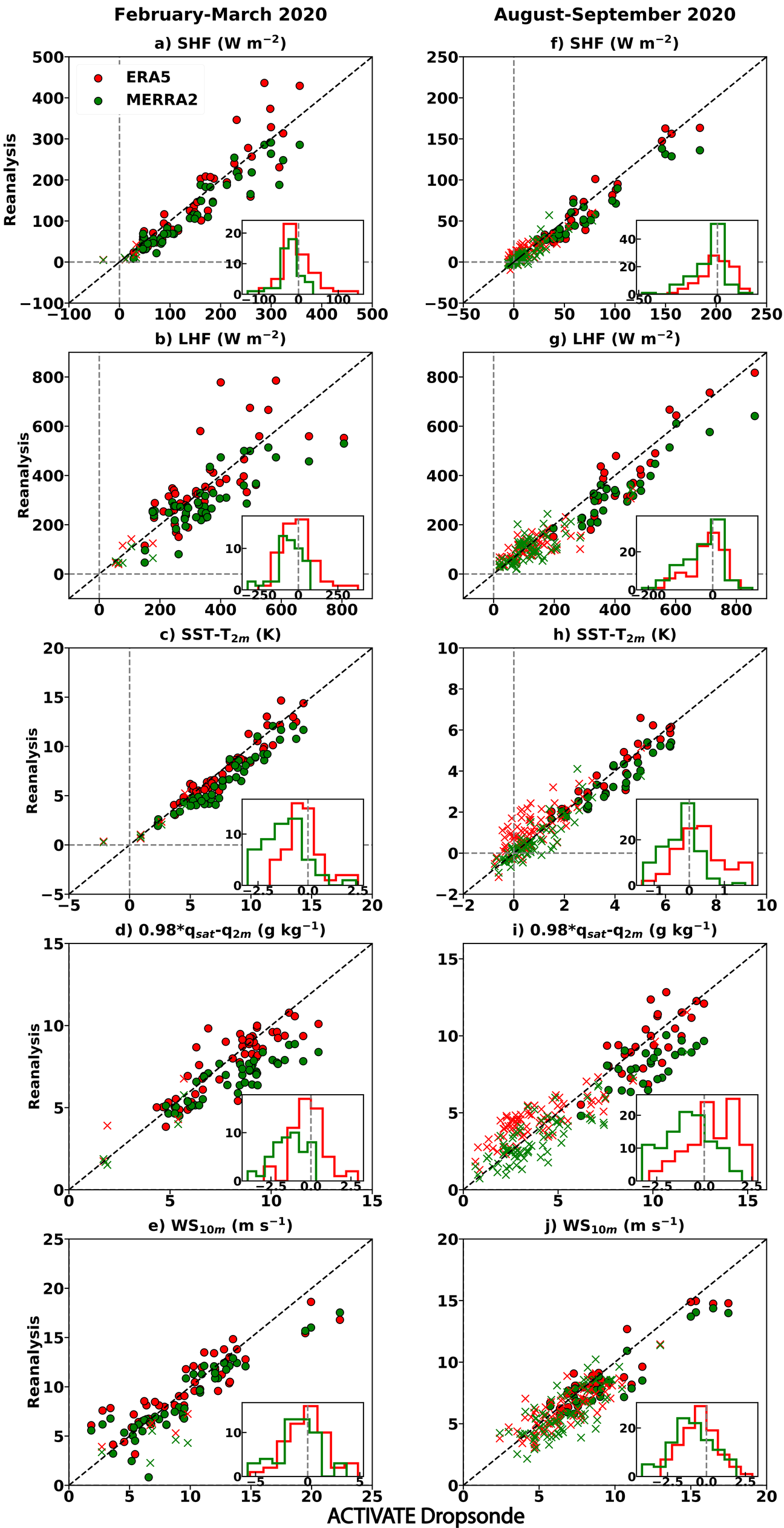
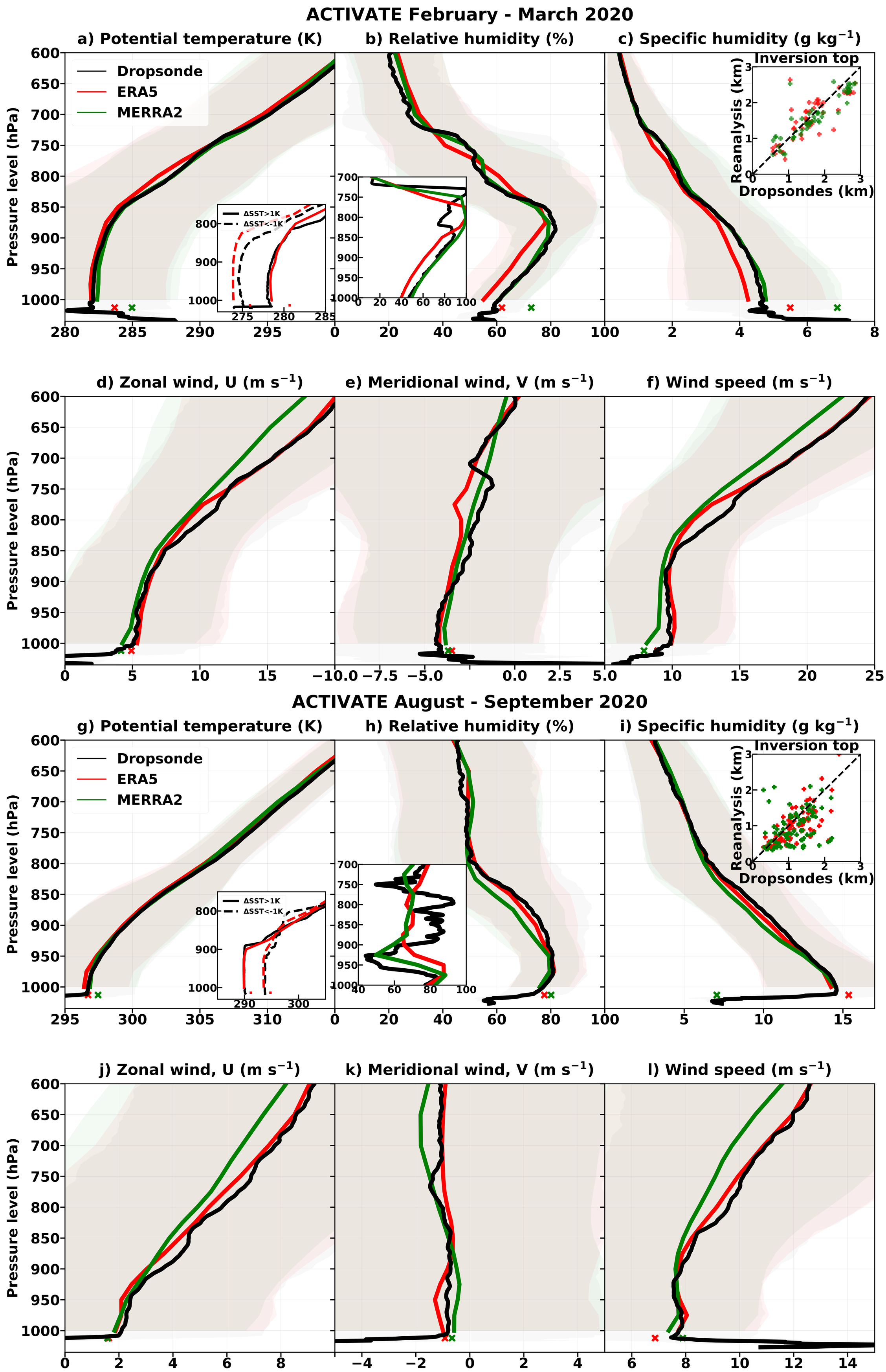

February-March 2020
August-September 2020
a) SHF (W m^{-2})
f) SHF (W m^{-2})
b) LHF (W m^{-2})
g) LHF (W m^{-2})
c) SST-T$_{2m}$ (K)
h) SST-T$_{2m}$ (K)
d) 0.98*q$_{sat}$-q$_{2m}$ (g kg^{-1})
i) 0.98*q$_{sat}$-q$_{2m}$ (g kg^{-1})
e) WS$_{10m}$ (m s^{-1})
j) WS$_{10m}$ (m s^{-1})
ERA5
MERRA2
Reanalysis
ACTIVATE Dropsonde

Figure 4.


ACTIVATE February - March 2020
a) Potential temperature (K)
b) Relative humidity (%)
c) Specific humidity (g kg^{-1})
Dropsonde
ERA5
MERRA2
ΔSST>1K
ΔSST<-1K
Inversion top
Reanalysis (km)
Dropsondes (km)
Pressure level (hPa)
d) Zonal wind, U (m s^{-1})
e) Meridional wind, V (m s^{-1})
f) Wind speed (m s^{-1})
ACTIVATE August - September 2020
g) Potential temperature (K)
h) Relative humidity (%)
i) Specific humidity (g kg^{-1})
j) Zonal wind, U (m s^{-1})
k) Meridional wind, V (m s^{-1})
l) Wind speed (m s^{-1})