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 Institude, 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 convection, in particular where the Gulf Stream intersects the mid-latitude storm track during the off-summer months. The westward-moving synoptic disturbances exchange warm, low-latitude air with the cold, high-latitude air that commonly trail the cold. 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 circulations characteristic of these clouds (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), with subsequent nudging for some 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 fluxes derived from aircraft data, which require long, low-altitude level legs that compete with other objectives such as 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 seeks to address 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 as our 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 are often 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

Buoy measurements of near-surface air variables and direct covariance buoyancy fluxes from the CLIMODE (Climate Variability and Predictability Mode Water Dynamics Experiment; Marshall et al., 2009) campaign provide absolute reference values against which to gauge the reanalyses. 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, the buoy was 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

123 the surface fluxes are responsive, and a sub-surface foundation SST (*SST_foundation*) are
124 measured and reported, as are the direct covariance buoyancy fluxes (Edson et al., 2013).

125 The ACTIVATE campaign sampled on both sides of the northern Gulf Stream SST
126 gradient (Fig. 1). During February-March, the Gulf Stream meandered to the north, more
127 noticeable west of 73°W, altering the local SST by more than 4K in places (Fig. 1). The more
128 quiescent synoptic conditions in August-September over a more uniform area support an
129 assessment of boundary-layer depictions less influenced by strong air-sea interactions.
130 Dropsondes were launched from the UC-12 King Air flying at approximately 9 km. Of the 13
131 King Air flight days in February-March, 2020, 8 coincided with visually-identified CAOs,
132 encompassing 43 of the 59 winter dropsondes (Table S1 lists the individual flight days and their
133 designation as CAO/non-CAO days). Dropsonde circles provide more intensive sampling of a
134 February 28 CAO over the warm sector of the Gulf Stream, and another on March 1 straddling a
135 sharp SST gradient to the north, and are the subject of detailed simulations (Li et al., 2021). The
136 18 August-September flight days include 3 (weaker) CAOs and deployed 107 dropsondes total**.**
137 Neither the buoy nor the dropsonde data are assimilated in either reanalysis or flux product.

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

141 The ERA5 Reanalysis is arguably the publicly-available reanalysis with the most
142 sophisticated depiction of the cloudy boundary layer. ERA5 has a horizontal grid spacing of 31
143 km, 137 vertical levels, and an hourly temporal resolution. A systematic bias in the partitioning
144 of ERA5's global wind kinetic energy, with an excessive mean zonal flow coupled with weak
145 meridional flow, is attributed to difficulty in representing high-frequency transient atmospheric
146 events (Rivas and Stoffelen, 2019). In addition, the mid-latitude boundary layer maintains a cold
147 and dry bias (Hersbach et al., 2020). An important aspect of ERA5 for this study is that the
148 assimilated SST product prior to 2007 was based on the Hadley Centre Sea Ice and Sea Surface
149 Temperature data set (HadISST2; Titchner & Rayner, 2014) but after 2007 it was based on the
150 Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA, Donlon et al., 2012).
151 HadISST2 is a 0.25°x0.25° pentad product developed from *in-situ* and infrared satellite data and
152 known to not resolve Gulf Stream SST gradients well (Chelton and Risien, 2016). OSTIA is
153 produced daily on a higher-resolution 0.05^{o}x0.05^{o} grid, and includes 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 differences in 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 (Gelaro et al., 2017) includes an aerosol climatology, which can potentially inform on the aerosol type contributing to CAO cloud characteristics (Painemal et al., 2021; Dadashazar et al., 2021). The MERRA2 spatial resolution is 0.625^o x 0.5^o 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, can be more easily combined with contemporaneous similarly-gridded satellite datasets (Painemal et al., 2021) and provides further insight into the impact of spatial resolution. OAFLUX integrates satellite-derived cloud data to realistically balance the surface fluxes (Yu et al., 2008). A bulk algorithm computes the surface latent and sensible heat fluxes using the Coupled Ocean-Atmosphere Response Experiment (COARE) v3.0 bulk parameterization (Fairall et al., 2003).

3. Method

Fluxes calculated using the COARE v3.0 bulk parameterization applied to the buoy T_{2m}, q_{2m} and WS_{10m} compare well against the buoy direct covariance buoyancy fluxes (Edson et al., 2013; Bigorre et al., 2013). The COARE v3.5 bulk flux algorithm incorporates modifications at wind speeds > 12 m s^{-1} based on the CLIMODE measurements (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 at the high wind speeds characteristic of CAOs. Surface fluxes are calculated from the individual dropsondes using the v3.5 parameterization. Dropsonde inter- or

extrapolation provides T_{2m} and q_{2m} and an *SST_skin* is estimated from GHRSST *SST_foundation* + (ERA5 *SST_skin* – ERA5 *SST_foundation*).

In the dropsonde comparison to MERRA2 profiles, 5 of the 58 MERRA2 winter near-surface temperature values exceeded measurements by 10K, were more moist by over 4 g kg^{-1}, and windier by up to 10 m s^{-1}. Such outlier behavior may reflect a mismatch in the timing of location of the cold front, but is also documented in Bosilovich et al. (2015). These comparisons are removed from this study, and such profiles could also be eliminated by evaluating neighboring MERRA2 profiles.

Instantaneous values captured by the dropsondes, from, e.g., cold pools and the up- and downdrafts within the CAO roll circulations will increase the variability of the dropsonde-calculated fluxes beyond that of the coarser-resolution (both spatially and temporally) reanalysis fluxes. We note the two dropsonde circles occurred over overcast clouds, prior to their breakup into clouds oriented along roll circulations. The reanalysis comparisons to the dropsondes occur nearest in time and space, while 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 background on the datasets and method 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. S3). 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 also 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 overestimation of >80%. Since the bulk flux calculations themselves have been previously 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. All of the products overestimate the SST (Fig. 2b), most notably the poorer spatial-resolution OAFLUX. This 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 ERA5 and MERRA2 match the buoy wind speeds well (Fig. 2f, Table S2). The surface q_s is elevated for all reanalyses/product, as expected. Interestingly, a dry bias in ERA5's q_{2m} contrasts with a moist bias for MERRA2 q_{2m}, both by about 1 g kg^{-1}. In combination, the air-sea temperature and humidity differences are smaller for MERRA2 on most (but not all) days, compared to ERA5 (Fig. 2g and h, Table S2), compensating for MERRA2's poorer near-surface air temperatures (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 SST and near-surface meteorology, likely in part because of a horizontal grid spacing that is approximately one-half that of MERRA2, but that compensations within MERRA2 model physics may be improving the buoyancy 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) by all three products, but those by ERA5 are the most pronounced. This is explained by air-sea temperature and humidity differences that are largest for ERA5, reflecting a too-cool T_{2m} and too-dry (by 1-2 g/kg) q_{2m}. The consistent too-dry ERA5 q_{2m} bias between the two seasons indicates a common bias source. In contrast, a too-warm T_{2m} in winter alternating with a too-cool T_{2m} in summer for ERA5, suggests the underlying causes may differ.

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 range in the surface fluxes estimated from the ACTIVATE dropsondes (not shown) are comparable to those from the CLIMODE buoy measurements, indicating that a fully representative set of synoptic conditions were sampled. 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 cannot be traced to pronounced air-sea temperature or humidity differences (Fig. 3c,d; Table S3), however. In contrast, the MERRA2 fluxes underestimate, at both high and low flux values. The underestimate is clearly linked to 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 *RH* and wind biases are smaller in August-September, consistent with a more homogeneous ocean surface temperature and weaker overall winds. 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 (after the exclusion of the 5 profiles) 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.

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 entirely constrain a full 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 acts to raise the boundary layer height, all else being equal. As discussed eloquently in Sandu et al., (2013), 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. 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 known difficulties, this study documents a wintertime ERA5 boundary layer depth, thermodynamic and dynamic structure that 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 boundary layer clouds, as the depth affects the coupling to the ocean surface supplying heat and moisture to the cloud layer. In addition, the robustness of the bias in ERA5 specific humidity suggests that one solution could be to adjust the ERA5 value by an observationally-determined correction factor, one that can be improved upon 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 dropsonde data coming available from further ACTIVATE deployments 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, with too-weak surface fluxes. Nevertheless the boundary layer depth depiction is similar: approximately realistic during the winter, and too shallow during the summer.

One consequence of the momentum transport to the surface by ERA5, which can increase the wind speed because of the overall increase of the winds with height, is that the boundary layer may deepen more quickly in ERA5, and encourage a faster cloud transition, than is seen in nature (Saggiaroto et al., 2020). Future work will incorporate space-based lidar and radar data along with ACTIVATE measurements to further document the evolution of the boundary layer and cloud structure of cold-air-outbreaks as they flow 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/. The TOGA-COARE v.3.5 parameterization is available from ftp://ftp1.esrl.noaa.gov/BLO/Air-Sea/bulkalg/cor3_5/python/.

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

Dadashazar, H., Painemal, D., Alipanah, M., Brunke, M., Seethala , C., Corral, A. F., et al., (2021). Cloud Drop Number Concentrations over the Western North Atlantic Ocean: Seasonal

Cycle, Aerosol Interrelationships, and Other Influential Factors. *Atmospheric Chemistry and Physics Discussions*, https://doi.org/10.5194/acp-2021-153.

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

Gelaro, R., McCarty,W., Suárez, M.J., Todling, R.,Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R.,Wargan, K., Coy, L., Cullather, R.,Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M. & Zhao, B. (2017). The Modern-Era Retrospective analysis for Research

and Applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. doi: 10.1175/JCLI-D-16-0758.1

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