



FINAL REPORT

Planetary Boundary Layer (PBL) Verification with 2022 TRACER-AQ Field Study Data

TCEQ Contract No. 582-23-45974
Work Order No. 2

Revision 2.0

Prepared by:

John Henderson, Archana Dayalu, and Matthew Alvarado
Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Ave.
Lexington, MA 02421
Correspondence to: adayalu@aer.com

Prepared for:

Bryce Kuchan
Texas Commission on Environmental Quality (TCEQ)
Air Quality Division
Building E, Room 383
Austin, Texas 78711-3087

26 June 2024

Document Change Record

Revision	Revision Date	Remarks
0.1	31 May 2024	Internal Version for Review
1.0	12 Jun 2024	Delivery to TCEQ
2.0	26 Jun 2024	Final Version Delivered to TCEQ

TABLE OF CONTENTS

Executive Summary	9
1 Introduction	10
1.1 Project Objectives.....	10
1.2 Background	11
1.3 Report Outline	11
2 Improvement to Radar Based PBLH Estimation Methodology	12
2.1 AMDAR “Ground Truth” Data: Overview and Caveats	12
2.2 Radar PBLH Algorithm improvement methodology using AMDAR.....	14
2.2.1 Step 1: Selection of Case Study Days.	15
2.2.2 Step 2: Identification of pervasive problems.	18
2.2.3 Step 3: Algorithm Modification.....	20
2.2.4 Step 4: Evaluation of Modifications	22
3 Validation and Evaluation of the radar-derived PBLH Estimation Methodology in Houston.....	31
3.1 Summary of TRACER-AQ Data	32
3.2 Results	33
4 Quality Assurance	36
5 Conclusions	36
6 Recommendations for Future Work.....	36
7 References	38
8 Appendix A: Time vs Height Radar Code Docker Instructions.....	40

List of Figures

Figure 1. AMDAR data locations across CONUS. Texas airports used in this work are circled in black. Figure adapted from Ayazpour et al. (2023). 13

Figure 2. Location offset between NEXRAD-AMDAR observation pair. Radar location is indicated by black dot and box surrounding it; ring of locations where radar beam intercepts the PBL top is shown by blue oval. Red airplane icon shows AMDAR (airport) location. The blue ring expands and contracts depending on depth of PBL. 13

Figure 3. EWX radar *zdr* field overlaid with radar PBLH trace (black circles), and AMDAR PBLH estimates (grey stars) on 27 May 2018. Flatlining of radar PBLH trace is evident..... 18

Figure 4. Illustration of Radar PBLH performance (HQ cases) compared to AMDAR, aggregated across all sites on HQ days from 2017-2019. RMA regression lines are provided for cases with and without two outliers removed. 19

Figure 5. Illustration of Radar PBLH performance (HQ cases). (a) comparison to AMDAR, aggregated across all sites on HQ days from 2017-2019. (b) Example of radar PBLH estimates (black points) and *zdr* field from NEXRAD site GRK tracking with AMDAR points (grey stars) from the BSM airport overlaid. 20

Figure 6. PBLH radar estimates versus AMDAR observations for all four fix experiments using GRK-BSM JJA as an example. Red line is RMA regression fit; black dashed line is 1:1. Radar PBLH algorithm performance across all 12Zfix experiments is consistent when compared to AMDAR PBLH estimates. Clockwise from top: 12ZfixA; 12ZfixC; 12ZfixCm1; and 12ZfixCo. 24

Figure 7. PBLH estimates using (left) 12ZfixCo and (right) 12ZfixCm1 at HGX-HOU for DJF season. Flatlining in 12ZfixCm1 is pronounced. Red line is RMA regression fit; black dashed line is 1:1. 25

Figure 8. Impact of 12ZfixCo modifications. Scatterplots of radar PBLH estimates vs AMDAR PBLH for each NEXRAD-AMDAR pair’s seasonal evaluation subset. Rows 1-4 exhibit results for EWX-SAT MAM, GRK-BSM JJA, FWS-DFW SON, and HGX-HOU DJF, respectively. Left column shows results using original Baseline code; right column shows 12ZfixCo results. 27

Figure 9. EWX *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from SAT on 27 May 2018. (Top) Reproduction of Baseline case as shown in Figure 3; (Bottom) Correction to flatlining behavior following “12ZfixCo” modifications. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic. 28

Figure 10. GRK *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from BSM on 18 Aug 2019. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic. 29

- Figure 11.** FWS *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from DFW on 3 Nov 2018. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic. 30
- Figure 12.** HGX *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from HOU on 29 Jan 2018. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic. 31
- Figure 13.** Map of radar and validation sites. Houston NEXRAD (HGX) radar location is indicated by black dot and box surrounding it; ring of locations where radar beam intercepts the PBL top is shown by blue oval and expands or contracts depending on PBLH depth. Red dot shows the AMDAR (airport) locations at HOU (left panel) and IAH (right panel). TRACER-AQ sonde release sites (M1, S1) are at red triangle. Note that for Houston area observations, Gulf influences were filtered and only PBL-radar beam intersections in the northwest quadrant were retained. 32
- Figure 14.** Boxplots of TRACER-AQ sonde-derived PBLH estimates using four methods on the 35 case study days. Mean difference among methods ranges from approximately 40 to 400m. 35
- Figure 15.** Mean *zdr* field for Houston HGX on 20220302 (elevation angle 5.1 degrees, azimuthal range 270-360 degrees). 12ZfixCo PBLH trace is indicated by black dots. TRACER-AQ PBLHSONDE1MCFARL estimates are provided by grey shapes, with each shape representing one of four estimation methods. The spread of sonde data estimates is indicated by the black circles. 35

List of Tables

Table 1. Schedule of Deliverables for Work Order No. 2 10

Table 2. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair GRK-BSM. 16

Table 3. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair FWS-DFW..... 16

Table 4. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pairs HGX-IAH and HGX-HOU. Note that Houston area airports share the same NEXRAD observations, and therefore have the same list of dates. 17

Table 5. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair EWX-SAT..... 17

Table 6. Summary of four experiments conducted for radar-PBLH corrections. The 12Z component of the experiment name refers to the need to address the starting time of the first PBLH estimates. Max z refers to the maximum PBLH allowed during the 1200 UTC hour. Change factor from prior (f, shortened here for space from maxgrowthfactor in the code) refers to a height-dependent multiplier of the prior-time height estimate that is subsequently modulated at different times by the Prescribed Growth to define a height range in which to search for the minimum in *zdr*. Lower *zdr* limit refers to the threshold below which data at that (height,time) pixel in all fields is removed. 21

Table 7. Average (SD) for AMDAR and five radar algorithm code versions for each site and season subset. MAM=Spring; SON=Fall; JJA=Summer; DJF=Winter..... 22

Table 8. Radar PBLH compared with AMDAR for five radar algorithm code versions. Mean Bias (MBE=AMDAR-radar), Root Mean Square Error (RMSE), and Pearson’s R are provided. 23

Table 9. Summary of all TRACER-AQ 2022 PBLH Data. TRACER-AQ Main Site (M1) refers to location at LaPorte, TX (29.67N, -95.059E); Supplementary Site (S1) is located at M1; Supplementary Site (S3) refers to location at Damon, TX (29.328N, -95.741E). Note that sites M1 and S1 are co-located and for the purposes of this study are treated as identical. ***Only data set used in validation exercise.* 33

Table 10. 39 dates used in TRACER-AQ comparison. Bolded dates showed HQ baseline algorithm performance against TRACER-AQ sonde data..... 33

Table 11. Comparison of Radar PBLH estimates with TRACER-AQ PBLHTSONDE1MCFARL dataset for case study (non HQ) days. Comparison with sonde estimates conducted before and after 12ZfixCo (as FixCo) modifications. Comparisons provided for each sonde PBLH method and as average aggregating all sonde estimates. Mean Bias Error (MBE, in meters); Root Mean Square Error (RMSE, in meters) and correlation coefficient (r) summarize base vs FixCo modifications. 34

Table 12. Comparison of Radar PBLH estimates with TRACER-AQ PBLHTSONDE1MCFARL dataset for 4 HQ days only. Comparison with sonde estimates conducted before and after 12ZfixCo (as FixCo) modifications.

Comparisons provided for each sonde PBLH method and as average aggregating all sonde estimates. Mean Bias Error (MBE, in meters); Root Mean Square Error (RMSE, in meters) summarize base vs FixCo modifications. Correlation coefficient (r) is not provided due to low sample size (n=4)..... 34

List of Acronyms

AMDAR – Aircraft Meteorological DATA Relay
AER – Atmospheric and Environmental Research
BSM – Austin-Bergstrom International Airport
CONUS – Contiguous United States
DFW – Dallas Fort Worth International Airport
DJF – December January February (Winter) Season
EPA – United States Environmental Protection Agency
EWX – San Antonio region NEXRAD
FCAA – Federal Clean Air Act
FWS – Dallas Fort Worth region NEXRAD
GRK – Austin region NEXRAD
HGX – Houston region NEXRAD
HOU – William P. Hobby Airport
HQ – High Quality data subset
IAH – George Bush Intercontinental Airport
JJA – June July August (Summer) Season
MAM – March April May (Spring) Season
MBE – Mean Bias Error
NEXRAD – Next Generation Weather Radar
NWS – National Weather Service
PBL – Planetary Boundary Layer
PBLH – Planetary Boundary Layer Height
QVP – Quasi Vertical Profile
R – Pearson Correlation Coefficient
RAOB – Rawinsonde Observation
RMA – Reduced Major Axis regression
RMSE – Root Mean Square Error
SAT – San Antonio International Airport
SD – Standard Deviation
SON – September October November (Fall) Season
TCEQ – Texas Commission on Environmental Quality
TRACER-AQ – Tracking Aerosol Convection interactions Experiment-Air Quality
UTC – Coordinated Universal Time
WRF – Weather Research and Forecasting Model
YYYYMMDD – Four-digit year, two-digit month, two-digit day date format
zdr – Differential Reflectivity

Executive Summary

The Planetary Boundary Layer Height (PBLH) is an important meteorological parameter for air quality concerns. The Planetary Boundary Layer (PBL) is defined as the region of the lower troposphere where meteorological conditions are largely impacted by the Earth's surface. The depth of this layer strongly influences air quality in a given region since it determines the volume of air into which pollutants are mixed.

In fiscal year (FY) 2022 and 2023, AER developed a radar-based PBLH estimation algorithm. The purpose of these radar PBLH estimates was to evaluate WRF model output. However, a major issue was the lack of “ground-truth” – or an independent data set – verification of the radar PBLH algorithm.

The purpose of this project was to refine the FY 2022 and 2023 scientific radar PBLH algorithm using newly available independent observations. To that end, AER used independent PBLH estimates derived from Texas-wide Aircraft Meteorological Data Relay (AMDAR) observations to provide a “ground truth” reference for adjusting and tuning the radar PBLH algorithm. PBLH algorithm tuning was based on AMDAR data from five Texas airports: Austin-Bergstrom International (formerly BSM), Dallas Fort-Worth International (DFW), William P. Hobby Airport Houston (HOU), George Bush Intercontinental (IAH), and San Antonio International (SAT). AER also used sonde-based PBLH estimates from the 2022 Tracking Aerosol Convection Experiment – Air Quality (TRACER-AQ) field study in Houston to further evaluate adjustments in the algorithm.

There were two major tasks to accomplish this project's objectives. The first of the tasks performed the actual AMDAR comparison to inform radar PBLH code modifications. This task involved updating the FY 2022 and 2023 software to have improved algorithms to use for the radar estimated PBLH methodology and evaluating the improvements with the AMDAR data set. All changes were made “under-the-hood” to keep TCEQ's usage of the original FY 2022 and 2023 software consistent. The second task extended the evaluation by comparing original and modified radar algorithm results from Houston with sonde-based PBLH estimates from the Houston TRACER-AQ 2022 campaign.

Overall, AER was able to find and troubleshoot key deficiencies in the original FY 2022 and 2023 radar-based algorithm. The modifications led to dramatic improvements in PBLH estimation when compared to AMDAR data for all non-Houston sites. The optimal radar algorithm version typically estimated 1.2-1.5 times greater than the AMDAR PBLH with significant increases in correlation and reductions in mean bias when compared to the original FY 2022 and 2023 algorithm. When compared to TRACER-AQ sonde-based estimates from Houston, the original radar algorithm significantly underestimated PBLH. In contrast, the optimal corrected algorithm tended to overestimate PBLH relative to the sonde data, but the mean bias was significantly reduced from the original algorithm. Also found that the sonde PBLH estimation methods themselves can vary over a considerable height range. Despite overall improvement, correlation remained poor likely due to low data availability from the Houston radar.

In the future, a recently developed daily data set of PBLH estimates (13:00-14:00 local solar time) over the contiguous US from 2005-2019 can be used in broader evaluations of the PBLH radar methodology. In addition, as sonde estimates in other Texas cities becomes available, future work should evaluate performance outside of the Houston area.

1 Introduction

1.1 Project Objectives

The purpose of this project is to improve and verify performance of the previously developed radar PBLH algorithm using Aircraft Meteorological Data Relay (AMDAR) observations and 2022 Tracking Aerosol Convection Experiment – Air Quality (TRACER-AQ) field study data.

The current project extends the work accomplished by AER in FY22 and FY23 where a radar-based PBLH algorithm was developed and used to evaluate WRF output (Henderson & Mountain, 2022; Henderson & Vagasky, 2023). As part of the current project, AER has (1) identified key deficiencies in the earlier version of the radar PBLH algorithm and has applied further improvements by constraining with hourly AMDAR-derived PBLH “ground truth” observations from 2017-2019 for five airports in the Texas region – BSM, DFW, HOU, IAH, and SAT; and (2) verified improvements by comparison with PBLH estimates from the Houston TRACER-AQ campaign.

The Schedule of Deliverables for this project is given in Table 1.

Table 1. Schedule of Deliverables for Work Order No. 2

Milestones	Planned Date
Task 1 - Work Plan	
1.1: TCEQ-approved Work Plan	February 12, 2024
1.2: TCEQ-approved QAPP	February 12, 2024
Task 2 – Progress Reports	
2.1: Monthly Progress Reports	Monthly
Task 3 – Improvement to Radar Based PBLH Estimation Methodology	
3.1: Updated Software and user manual	April 30, 2024
Task 4 – Validation and Evaluation of the radar-derived PBLH Estimation Methodology	
4.1: Technical memo describing results from the validation of the updated radar-derived PBLH estimates for 2022	May 31, 2024
Task 5 – Training Presentation for TCEQ Air Modeling and Data Analysis Staff	
5.1: Virtual, recorded training presentation for TCEQ air modeling and data analysis staff	May 31 2024
Task 6 – Draft and Final Reports	
6.1: Draft Report	June 14, 2024
6.2: Final Report	June 28, 2024
6.3: Docker container of final version of software package and user manual	June 28, 2024

1.2 Background

TCEQ is required under the Federal Clean Air Act (FCAA) to perform air quality modeling for attainment demonstration purposes. Air quality modeling includes simulations of meteorology using the Weather Research and Forecasting (WRF) model. Evaluating model performance is a crucial step in the air quality modeling process. Model performance involves comparing simulated values of a certain parameter to estimated or measured values of the same parameter and quantifying the difference using various statistical measures. Model performance is often a multi-directional analysis and involves using as many comparisons as possible.

The Planetary Boundary Layer Height (PBLH) is an important meteorological parameter for air quality concerns. The Planetary Boundary Layer (PBL) is defined as the region of the lower troposphere where meteorological conditions are largely impacted by the Earth's surface. The depth of this layer strongly influences air quality in a given region since it determines the volume of air into which pollutants are mixed. Banghoff et al. (2018) showed a method to use the differential reflectivity field of National Weather Service (NWS) radars to estimate PBLH under certain weather conditions, which can then be used to validate WRF model simulations.

As part of previous contract work, AER developed software for determining average PBLH from 88D weather radars. This method provides an average PBLH over the area sampled by the radar, and thus does not represent a single point measurement. AER also developed an operator to compare this PBLH estimate to WRF model output from TCEQ's 2019 modeling platform. The current work sought to improve the radar-derived PBLH estimation method by validating it against previously unavailable PBLH "ground truth" observations derived from Aircraft Meteorological Data Relay (AMDAR) data spanning 2017-2019 (Li et al., 2020). Li et al. (2020) provides AMDAR PBLH derivations for five airports in the Texas region – Austin-Bergstrom International (formerly BSM), Dallas Fort-Worth International (DFW), William P. Hobby Airport Houston (HOU), George Bush Intercontinental (IAH), and San Antonio International (SAT). AMDAR data, with its incorporation of lower atmosphere Rawinsonde Observation (RAOB) profiles, provides the sufficiently high temporal frequency needed to represent the PBL, and provides valuable improvements to the existing radar algorithm. In addition, TRACER-AQ sonde observations from Houston for the 2022 measurement period were used to evaluate the AMDAR-based improvements to the PBLH radar technique.

1.3 Report Outline

This Final Report highlights major activities and key findings, provides pertinent analysis, describes encountered problems and associated corrective actions, and details relevant statistics including data, parameter, or model completeness, accuracy and precision. This report satisfies Deliverable 6.2 of the Work Plan for Work Order No. 2 under TCEQ Contract 582-23-45974:

Deliverable 6.2: Final Report

Deliverable 6.2 Due Date: June 28, 2024

2 Improvement to Radar Based PBLH Estimation Methodology

Banghoff et al. (2018) presented a method to use the differential reflectivity field of NWS radars to estimate PBLH. Differential reflectivity (zdr) is a Level 3 product from NWS 88D weather radars that became available following the upgrade in recent years to dual polarization. This field has the potential to estimate PBLH under certain weather conditions, which can then be used to validate model simulations and thus ameliorate the systematic lack of PBLH observations. In Henderson & Mountain (2022), AER developed and installed a proof-of-concept system for TCEQ that applied the quasi-vertical profile (QVP) methodology used by Banghoff et al. (2018) to NWS 88D weather radars at selected sites in Texas. The Henderson & Mountain (2022) radar based PBL heights were validated with an ozonesonde dataset from the Houston area provided by the TCEQ.

However, there were some key issues identified by Henderson & Mountain (2022) and follow-on WRF validation work by Henderson & Vagasky (2023) that motivated continued algorithm development. First, the small number of days in the Henderson & Mountain (2022) ozonesonde validation dataset, combined with the prevalence of non-optimal meteorological conditions on those days, prevented a robust quantitative assessment of the PBLH algorithm. Second, systematic issues were present ranging from unwanted influence from the residual layer to unphysical PBL growth and decay parameterization that was not mirrored in the corresponding zdr fields.

Fortunately, the recent availability of AMDAR-derived estimates of PBLH from five Texas area airports (Ayazpour et al., 2023; Li et al., 2020) presented the opportunity for a thorough validation of the radar PBLH algorithm and resolution of its deficiencies. The remainder of this section describes the AMDAR validation data set, the methodology to further develop the radar PBLH algorithm, and the impacts on model performance.

2.1 AMDAR “Ground Truth” Data: Overview and Caveats

Improvements to the radar algorithm PBLH estimates were made in part based on comparison to those from AMDAR observations (Zhang et al., 2020; Li et al., 2020). The raw AMDAR meteorological fields were processed into PBLH estimates by Li et al. (2020) using vertical profiles from aircraft at specific airports in the continental United States (Figure 1). For Texas, these included five airports: Austin-Bergstrom International (formerly BSM), Dallas Fort-Worth International (DFW), William P. Hobby Airport Houston (HOU), George Bush Intercontinental (IAH), and San Antonio International (SAT) (Figures 1 and 2). The raw meteorological observations that are processed into PBLH are aggregated for multiple aircraft taking off and landing at these airports. As such, the AMDAR PBLH estimates should be considered a time-average (over the 30 minutes prior to and following the observation time) and a spatial average based on the unique pattern of aircraft locations. The conversion from raw AMDAR meteorological observations into PBLH applies the widely used bulk Richardson method. This approach may not identify the same top to the PBL as the radar and is itself sensitive to the choice of internal parameters. Figure 2 shows the positioning of the five airports compared to the radar sites. Note that the radius of the ring of radar observations is dependent on the height of the boundary layer, with early morning time periods having a smaller radius corresponding to lower PBLH heights.



Figure 1. AMDAR data locations across CONUS. Texas airports used in this work are circled in black. Figure adapted from Ayazpour et al. (2023).

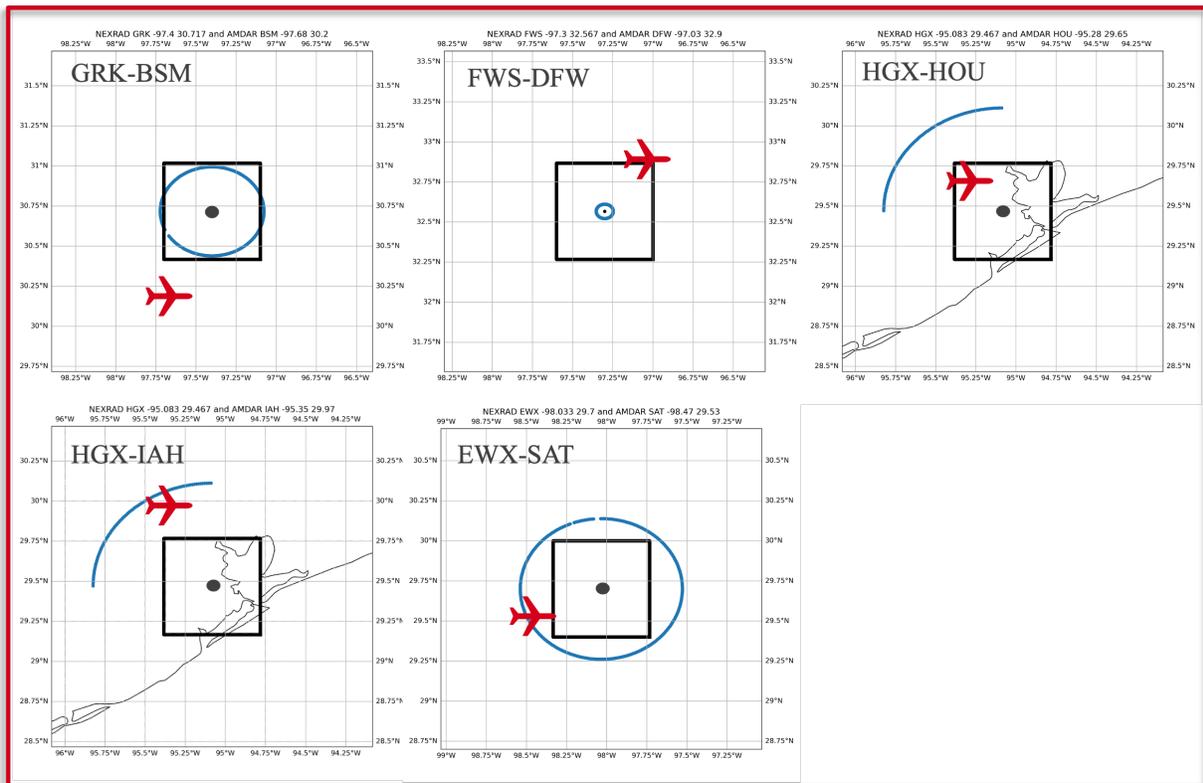


Figure 2. Location offset between NEXRAD-AMDAR observation pair. Radar location is indicated by black dot and box surrounding it; ring of locations where radar beam intercepts the PBL top is shown by blue oval. Red airplane icon shows AMDAR (airport) location. The blue ring expands and contracts depending on depth of PBL.

While hourly PBLH estimates are available from 2007 to October 2019, the most complete set of data are from 2017-2019. As there are no quality control flags present in

the dataset, all observations were used and weighted equivalently. Despite the categorization of the AMDAR PBLH data set as “ground truth”, it should be noted that the data are derived using the Bulk Richardson method from raw observations of the atmosphere and are estimates with their own degrees of uncertainty. From this perspective, not only do the raw observations and the subsequent AMDAR-derived PBLH have uncertainty, but the atmospheric boundary layer can also be poorly defined, especially at night. The AMDAR data thus has lower reliability prior to 1200 UTC – which is the starting time of application of the radar algorithm – and also in the early morning hours adjacent to 1200 UTC. During this time, the residual layers – i.e., remnants aloft of the prior day’s surface-based boundary layer – are present and often the AMDAR dataset erroneously identifies this as the current boundary layer. Thus, caution should be applied in interpreting the AMDAR PBLH estimates. In identifying the science code modifications for the experiments that follow, AER attempted to place the radar PBLH according to the *zdr* field, with some deference to the AMDAR location.

2.2 Radar PBLH Algorithm improvement methodology using AMDAR

A series of experiments were designed to improve the performance of the radar algorithm through both a statistical evaluation and a subjective evaluation. The latter involves inspection by a human to identify their preferred placement of the PBL top based on experience and supplemental data like PBLH observations and surface weather reports. At a top level, work performed in 2024 proceeded with the following four steps:

- **Step 1: Identifying a set of representative case study days** by running the original code daily. Case study days were those that exhibited a trough in the radar representation but that also did not perform optimally. “Optimally” here refers to situations where the radar algorithm should be correctable so that the PBLH estimate is positioned in the center of the trough in the *zdr* field and is in reasonable agreement with validation datasets, such as AMDAR PBLH observations. Addressing this need is a necessary step towards improving the usability of the overall radar approach and represents the first major effort towards addressing the recommended “Future Work” proposed in prior years (Henderson & Mountain, 2022; Henderson & Vagasky, 2023).
- **Step 2: Identifying pervasive problems** in the radar algorithm’s behavior on those case study days that would guide the specific scientific improvements to the algorithm.
- **Step 3: Modifying the algorithm.** This involved several iterations whereby the science code was first modified in an attempt to improve the radar placement of the PBLH within the *zdr* trough on a few selected case study days. Then, if the subjective evaluation of the changes was promising, a larger subset of the case study days was run through the modified code and validated in aggregate against AMDAR-derived PBLH observations. This approach resulted in a series of candidate code improvements.
- **Step 4: Evaluating the modifications** against baseline case study date subsets and an independent set of 2022 TRACER-AQ PLBH estimates. The TRACER-AQ results are discussed in Section 3.

Details of Steps 1 through 3 are documented in a User Guide provided to the TCEQ as Henderson (2024a). Step 4 is additionally documented in a Technical Memo provided to

the TCEQ as Henderson (2024b). The key components of each step are summarized below.

2.2.1 Step 1: Selection of Case Study Days.

AER selected 20 case study days per season for each NEXRAD radar site and AMDAR airport pair (GRK-BSM, FWS-DFW, HGX-IAH, HGX-HOU, and EWX-SAT) (Tables 2-5). The case studies were selected based on days that exhibited acceptable weather, a sufficiently clear rising trough in the *zdr* field representative of the growing convective boundary layer, and a radar-based PBLH trace that was imperfect but appeared correctable (e.g., Figure 3). To obtain the case study dates, AER ran the Henderson & Vagasky (2023) (hereafter “Baseline”) code each day separately in reverse chronological order for two years for each radar site/airport combination starting with the most recent day with AMDAR observations (i.e., YYYYMMDD = 20191031) and manually inspected each day’s *zdr* field, radar-derived PBLH estimates, and AMDAR PBLH observations. This took advantage of the higher data count of daily observations for more recent years in the long-period AMDAR dataset.

The Baseline case study provided a rich and varied dataset of the myriad ways that the *zdr* field can evolve in the presence of clouds, precipitation, and clear insect contaminations. The backdrop of these daily variations is the pattern of ground clutter and geographical nuances associated with each site. The baseline code was run using all azimuth angles except for Houston for which the scanning angles were restricted to the NW quadrant in an attempt to minimize the influence of the marine boundary layer over the Gulf of Mexico.

It was necessary to consider a balance between the importance of the radar representation and the placement of AMDAR observations. As noted in Section 4, all datasets have sources of error. There were many time periods (within each day) when the radar trough did not agree with the AMDAR PBLH estimate. This is likely a combination of a poorly defined PBL and top, as well as the presence of residual layers. Surface weather observations were also inspected to ensure that days with pervasive thick cloud and rain were not used. On those days, a rising trough in the *zdr* field typically was not apparent, so those days largely were not selected as candidates. Perfectly clear days were infrequent, especially at some times of the year, so marginal cases – which were deemed important enough to accommodate – were included. These often proved difficult to correct. AER also selected seemingly “High Quality” (HQ) days (67 in total) for which the radar algorithm identified the PBLH in relation to the *zdr* trough very well in the Baseline code (Section 2.2.2).

Table 2. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair GRK-BSM.

DJF-Winter	MAM-Spring	JJA-Summer	SON-Fall
20190227	20190427	20190829	20191031
20190213	20190426	20190820	20191027
20190204	20190425	20190819	20191019
20190128	20190420	20190818	20191017
20190112	20190419	20190817	20191009
20181227	20190404	20190809	20191002
20181223	20190328	20190720	20190925
20181215	20190326	20190719	20190923
20181203	20190318	20190718	20190922
20180209	20190317	20190717	20190915
20180208	20180530	20190710	20190907
20180201	20180525	20190709	20181128
20180130	20180516	20190708	20181126
20180129	20180515	20190706	20181123
20180123	20180510	20190704	20181117
20180109	20180507	20190629	20181004
20171217	20180506	20190619	20180919
20171214	20180505	20190615	20180916
20171209	20180428	20190604	20180915
20171202	20180426	20180814	20180914

Table 3. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair FWS-DFW.

DJF-Winter	MAM-Spring	JJA-Summer	SON-Fall
20190225	20190525	20190830	20191027
20190205	20190506	20190829	20190927
20181202	20190505	20190826	20190920
20181201	20190504	20190822	20190917
20180129	20190426	20190819	20190908
20180125	20190421	20190817	20181117
20180124	20180529	20190810	20181104
20180122	20180528	20190721	20181103
20180114	20180523	20190717	20181029
20180110	20180522	20190706	20181028
20180108	20180518	20190702	20181006
20180106	20180514	20190628	20181005
20180105	20180512	20190627	20181004
20180103	20180511	20190625	20181003
20171221	20180506	20190621	20180902
20171215	20180505	20190608	20171124
20171214	20180429	20190603	20171122
20171213	20180427	20180830	20171121
20171211	20180426	20180826	20171120
20171210	20180310	20180805	20171102

Table 4. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pairs HGX-IAH and HGX-HOU. Note that Houston area airports share the same NEXRAD observations, and therefore have the same list of dates.

DJF-Winter	MAM-Spring	JJA-Summer	SON-Fall
20190218	20190515	20190710	20191027
20190213	20190505	20190628	20191019
20190212	20190504	20190619	20190925
20190130	20190426	20190614	20181126
20190111	20190425	20190613	20181124
20190109	20190419	20180813	20181120
20190106	20190415	20180728	20181116
20190105	20190326	20180722	20181115
20190104	20190324	20180721	20181114
20181222	20190321	20180720	20181102
20181218	20180518	20180719	20181029
20181204	20180508	20180718	20181028
20181203	20180506	20180717	20181021
20180226	20180430	20180716	20181014
20180129	20180428	20180710	20180919
20180123	20180424	20180702	20171122
20180114	20180423	20180701	20171118
20171220	20180419	20180630	20171107
20171215	20180416	20180627	20171106
20171209	20180415	20180602	20171102

Table 5. List of Baseline Case Dates as YYYYMMDD broken down by season for NEXRAD-AMDAR pair EWX-SAT.

DJF-Winter	MAM-Spring	JJA-Summer	SON-Fall
20190217	20190528	20190826	20191027
20190214	20190522	20190823	20191017
20190213	20190504	20190818	20191004
20190128	20190421	20190817	20190926
20190112	20190415	20190813	20190925
20181227	20190327	20190812	20190919
20181223	20190326	20190809	20190915
20181220	20190308	20190808	20190908
20181213	20180531	20190807	20190905
20181201	20180529	20190730	20190904
20180204	20180528	20190728	20181126
20180122	20180527	20190718	20181120
20180108	20180526	20190708	20181101
20180104	20180510	20190629	20181028
20171220	20180509	20190621	20181027
20171214	20180429	20190620	20181020
20171213	20180428	20190619	20181010
20171212	20180426	20190616	20180917
20171204	20180425	20190615	20171127
20171202	20180419	20190601	20171106

2.2.2 Step 2: Identification of pervasive problems.

The Baseline case study set of days provided a rich and varied dataset of the myriad ways that the *zdr* field can evolve in the presence of clouds, precipitation and clear-air insect contaminants. The backdrop of these daily variations is the pattern of ground clutter and influence of geographical features unique to each site. AER ran the Baseline code using all azimuth angles except for Houston for which the scanning angles were restricted to the NW quadrant to minimize the influence of the marine boundary layer over the Gulf of Mexico.

Manual inspection of each for the two-year period for each radar site allowed for characterization of the Baseline radar algorithm's performance across all seasons at the five NEXRAD-AMДАР sites across Texas. The Baseline radar algorithm's behavioral problems can be summarized as follows:

- i. The PBLH at the starting time of the application of the radar algorithm (1200 UTC) is frequently incorrect. The most common source of error is placement within the residual layer, which is not preferred since the focus was on the influence of the surface on the convective boundary layer growth during the morning and afternoon hours.
- ii. There are frequent wild gyrations of the PBLH estimate due to non-optimal application of an existing growth/decay restriction between adjacent radar times (typically 5-10 minutes).
- iii. A number of days exhibited a 'flatlining' of the PBLH estimates in which they do not rise in conjunction with the increasing depth of the boundary layer – instead remaining flat over time in an unphysical manner.
- iv. Some sites exhibited considerable missing data, potentially exacerbated in certain seasons. This was common near the surface and likely associated with side lobes of the radar and subsequent unphysical ground clutter, which is aggressively removed in the current code. At times, there were also prominent *zdr* troughs that appeared otherwise physical and usable, except that the centers were devoid of data. While not possible within the time constraints of this current work, in-depth analysis of seasonality in representation will be helpful for future analyses.

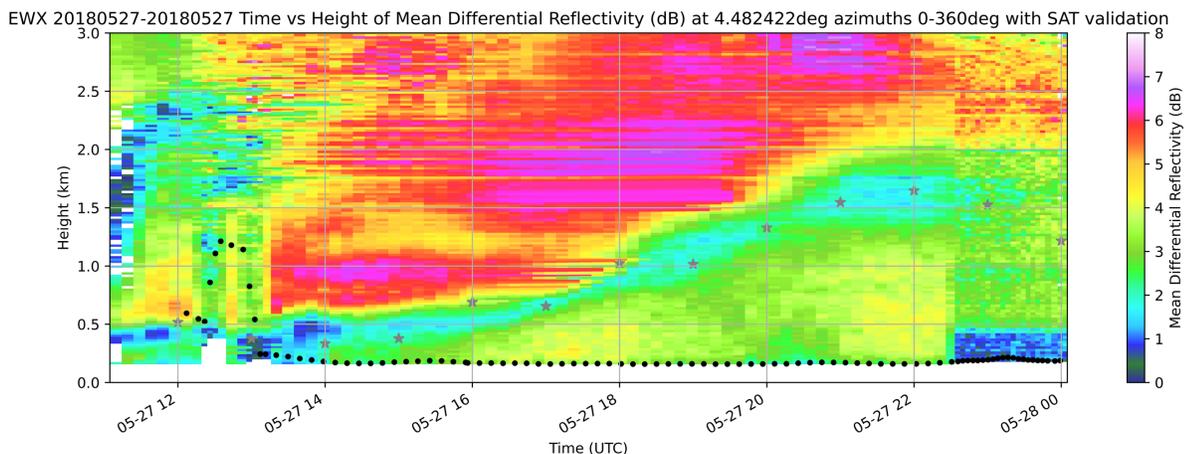


Figure 3. EWX radar *zdr* field overlaid with radar PBLH trace (black circles), and AMDAR PBLH estimates (grey stars) on 27 May 2018. Flatlining of radar PBLH trace is evident.

Figure 3 provides an example of all these deficiencies on a single day (EWX radar on 27 May 2018). The prominent rising trough in the zdr is marked by the steadily rising AMDAR PBLH estimates at SAT. The radar-derived PBLH trace (black circles) starts at 1200 UTC outside any discernable zdr features related to either the residual layer or a growing surface-based boundary layer. It then rises to approximately 1200 m then falls back near the surface and remains at the surface for the rest of the day. The initially high starting height, rapid fluctuations, and flatlining are common behavior seen on numerous other days. Also apparent on this plot is the small area of missing data at the surface between 1200 and 1300 UTC resulting from quality control filtering. On other days, this infrequently extends directly into the center of the rising trough.

For reference, AER also obtained a subset of dates for which the Baseline algorithm performed well. The purpose was two-fold: (1) examine the comparability of PBLH estimation methods, i.e., the extent to which AMDAR Bulk-Richardson PBLH estimates correlated with high-performing zdr -based PBLH estimates; and (2) retain this subset to evaluate the extent to which performance on these dates was robust to modifications to the Baseline algorithm. AER called these dates “high-quality” (HQ) dates. Of the two years inspected for four radar sites (365 daily plots for two years, totaling 2920 cases), only 67 were deemed to be performing well via the Baseline radar algorithm.

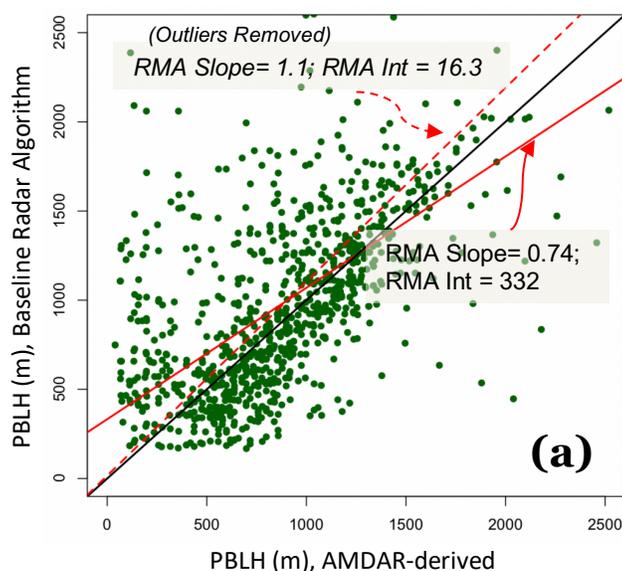


Figure 4. Illustration of Radar PBLH performance (HQ cases) compared to AMDAR, aggregated across all sites on HQ days from 2017-2019. RMA regression lines are provided for cases with and without two outliers removed.

Scatterplots of the HQ days (all sites and times are combined because of the small sample size) are shown in Figure 4. Reduced Major Axis (RMA) regression was used to account for uncertainties in both x (AMDAR PBLH estimates) and y (radar PBLH estimates) axes. On these days, the radar performed very well but not necessarily in the center of the zdr trough. Some cases were complex in nature, yet the radar algorithm excelled. Overall, there was good agreement with the HQ PBLH and the AMDAR PBLH; at times, the AMDAR observations were considerably offset from the zdr trough. Figure 4 presents linear regression statistics with and without outliers removed. The good

performance on these days demonstrates the potential of the radar algorithm. See Figure 5 for the Quasi Vertical Profile (QVP) plot for GRK for 8 May 2018. The radar trace is centered in the middle of the trough with few departures from what a trained meteorologist would consider as the likely top of the PBL based primarily on the zdr trace but also generally supported by the AMDAR observations. However, on this day, the AMDAR estimates were too high through 1500 UTC. The reader is reminded that the specific criteria by which the radar (height of minimum zdr) and AMDAR (bulk Richardson number) define the PBLH are different and so there are inherent biases in both estimates.

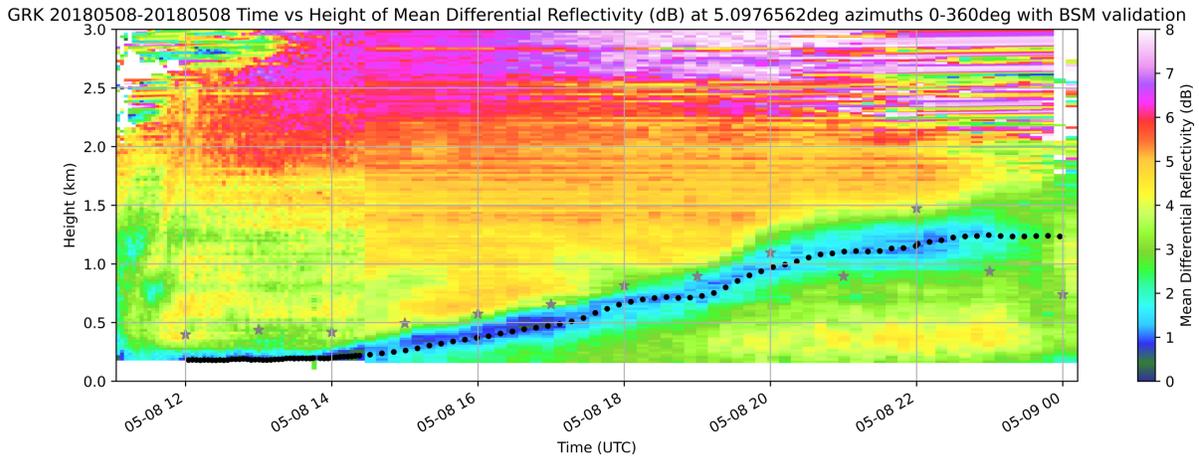


Figure 5. Illustration of Radar PBLH performance (HQ cases). (a) comparison to AMDAR, aggregated across all sites on HQ days from 2017-2019. (b) Example of radar PBLH estimates (black points) and zdr field from NEXRAD site GRK tracking with AMDAR points (grey stars) from the BSM airport overlaid.

2.2.3 Step 3: Algorithm Modification

The source of the four performance deficiencies described in Section 2.2.2 was investigated in the science code and a set of target code modifications was identified:

- i. Limit the starting height. This prevents the code from considering zdr minima well above the surface in the residual layer and from other non-physical artifacts. *This helps address the erroneously high initial PBLH issue.*
- ii. Enable the additional capacity of the “maxgrowthfactor” parameter being used to encourage PBLH growth at specific times of the day, instead of only being a means to prevent unphysical increases and decreases in PBLH estimates. *Along with (iii) below, this helps reduce unphysical instability in PBLH determination by limiting a height range in which to search for the local zdr minimum.*
- iii. Encourage searching for the growing PBL at levels above the prior estimate based on observed zdr features. *Along with (ii) above, this helps reduce unphysical instability in PBLH determination by limiting a height range in which to search for the local zdr minimum.*
- iv. Adjust the quality control filtering of the zdr field to allow more valid zdr pixels near the surface and in the zdr trough. The effect of this change varies

considerably across season and site, with the influence considerably greater for days with a weak *zdr* trough, ground clutter, and deep *zdr* troughs. *This helps address the data missingness issue.*

The science changes to the radar algorithm are designed to improve the overall performance and utility of the technique across all sites and seasons, including days for which the *zdr* representation is marginal. AER's code modifications and validation of these changes balance scientific simplicity and the need to perform within the time constraints of the project. Recommendations for future modifications are noted later in this document. Importantly, care was taken to not overfit individual cases by accommodating the marginal cases at the expense of the aggregate performance. It should be noted that the PBLH estimates from the radar are extremely sensitive to the chosen parameters, often making the difference between staying in the center of the *zdr* trough, to randomly rising out of it in an unphysical manner or, more common, 'flatlining'. The former situation can occur at any time; the latter situation occurs more frequently in the presence of clouds which can complicate the *zdr* representation via delayed and/or weakened surface heating. The results suggest major deficiencies of the code have been addressed in an easy-to-understand manner. In addition, appreciable subjective and objective improvements have been implemented such that the user can perform future adjustments if more specific concerns arise.

Table 6. Summary of four experiments conducted for radar-PBLH corrections. The 12Z component of the experiment name refers to the need to address the starting time of the first PBLH estimates. Max z refers to the maximum PBLH allowed during the 1200 UTC hour. Change factor from prior (f, shortened here for space from maxgrowthfactor in the code) refers to a height-dependent multiplier of the prior-time height estimate that is subsequently modulated at different times by the Prescribed Growth to define a height range in which to search for the minimum in *zdr*. Lower *zdr* limit refers to the threshold below which data at that (height,time) pixel in all fields is removed.

Experiment	Max z < 1300 UTC	Change factor from prior (f)	Prescribed Growth	Lower <i>zdr</i> Limit
Baseline (2023)	< 2000 m	1.1	None	0 dB
12ZfixA	< 300 m	z < 500: 1.5 z < 1000: 1.2 z >= 1000: 1.1	hour <= 17: [-z*0.5f,+z*f] +50 hour > 17: [-z*f,+z*f]	0 dB
12ZfixC	< 300 m	z < 500: 1.4 z < 1000: 1.2 z >= 1000: 1.1	hour <= 20: [-z*0.5f,+z*f] +25 hour > 20: [-z*f,+z*f]	-2 dB
12ZfixCo	< 300 m	As for 12ZfixC	As for 12ZfixC	0 dB
12ZfixCm1	< 300 m	As for 12ZfixC	As for 12ZfixC	-1 dB

As outlined above, the code was first modified to address the initial starting height concern for a few select days. When improvements were seen by running the code for these days and evaluating subjectively, for computational efficiency AER ran the

experiment configuration for a subset of cases in Tables 2-5. The subsequent process involved (1) selecting four site-season pairs (EWX-SAT MAM, FWS-DFW SON, GRK-BSM JJA and HGX-HOU DJF) out of the 16 total available in the baseline case studies; (2) performing manual inspection; and (3) regenerating objective statistics. This was repeated for four main experiments (Table 6).

The nomenclature of the experiment is as follows. The “12Zfix” component refers to the need to address the starting time of the first radar PBLH estimate. The majority of days had usable data at this 1200 UTC time. The next letter indicates the version of vertical height range permitted within the code to search for the trough in the *zdr* field (here, “A” and “C”) and reflects the growth of the convective PBL in the code. Experiment “B” was deemed unsatisfactory and not pursued. The trailing number, if present, represents the threshold below which data at that (height, time) pixel in all fields is removed if the *zdr* value at that pixel is below the threshold. The default threshold value is 0 dB in the 2023 Baseline code and “fixA”, while it is -2 dB in “fixC”, reset back to 0 dB for “Co” and an intermediate value of -1 dB for experiment “Cm1”. This threshold is an attempt to vary the amount of very low *zdr* values that are used by the radar algorithm. In all cases, manual inspection indicated that use of the values above did not result in inappropriate application of the approach.

2.2.4 Step 4: Evaluation of Modifications

Each experiment described in Table 6 was run for a subset of the original 80 case study days per site for computational efficiency. One season (=20 cases of days) was chosen for each site for time and efficiency: MAM for EWX-SAT; SON for FWS-DFW; JJA for GRK-BSM; and DJF for HGX-HOU. Given project constraints, this sampling was deemed sufficient for both manual subjective and objective statistical evaluation purposes. While objective validation is performed against the AMDAR dataset, keep in mind that the radar algorithm is being encouraged towards finding the *zdr* trough in all experiments, as per the scientific rationale underlying the technique. The manual inspection is used to explain the objective results.

Average PBLH and standard deviations (SD) from the 12UTC-23 UTC hour for each site and season are provided in Table 7. Averages and SD are provided for each of the AMDAR, Baseline, and four “12Zfix” experiments.

Table 7. Average (SD) for AMDAR and five radar algorithm code versions for each site and season subset. MAM=Spring; SON=Fall; JJA=Summer; DJF=Winter.

	EWX-SAT-MAM	FWS-DFW-SON	GRK-BSM-JJA	HGX-HOU-DJF
AMDAR	921.7 (488)	746.3 (429)	952.5 (451)	544.2 (312)
Radar: Baseline	1154 (791)	967.1 (589)	794.6 (781)	600.4 (372)
Radar: 12ZfixA	926.2 (562.5)	1125 (634)	1130 (589)	954.2 (635)
Radar: 12ZfixC	924.2 (643.4)	914.7 (664)	1166 (644)	474.9 (690)
Radar: 12ZfixCo	923.6 (621.7)	1030 (635)	1151 (621)	849.8 (660)
Radar: 12ZfixCm1	923.7 (641.3)	931 (629)	1199 (764)	566.5 (713)

Table 8. Radar PBLH compared with AMDAR for five radar algorithm code versions. Mean Bias (MBE=AMDAR-radar), Root Mean Square Error (RMSE), and Pearson's R are provided.

Site (Season)	MBE (m)	RMSE (m)	R	Version	Comments
EWX-SAT (MAM)	-214	964	-0.08	Baseline	Pervasive improper starting height, fluctuations, flatlining, joining trough later in day
FWS-DFW (SON)	-224	648	0.31	Baseline	Occasional, good performance but frequent missing filtered data; other problems common
GRK-BSM (JJA)	157	987	-0.19	Baseline	Pervasive typical problems at times with PBLH trace crossing trough at 90 degrees
HGX-HOU (DJF)	-58.8	494	-0.03	Baseline	Prominent residual layer at times, and occasional missing surface data; only a few captures; poor <i>zdr</i> signal often.
EWX-SAT (MAM)	3.45	401	0.59	12ZfixA	Generally, very good; some flatlines
FWS-DFW (SON)	-376	623	0.60	12ZfixA	Very good, some erratic behavior and flatlining with additional surface data
GRK-BSM (JJA)	-177	498	0.60	12ZfixA	Generally, very good with limited erratic behavior and a bit noisy
HGX-HOU (DJF)	-413	758	0.26	12ZfixA	Overall, very poor performance: flatlining on 17 days, with one favorable trough and two erratic days
EWX-SAT (MAM)	15.1	474	0.59	12ZfixC	Generally, very good but some flatlines
FWS-DFW (SON)	-165	587	0.49	12ZfixC	Very good, but some erratic behavior and flatlining in trough with additional surface data
GRK-BSM (JJA)	-213	530	0.61	12ZfixC	Generally, very good with limited erratic behavior and a bit noisy
HGX-HOU (DJF)	66.8	688	0.24	12ZfixC	Overall, very poor performance: flatlining on 17 days, with one favorable trough and two erratic days
EWX-SAT (MAM)	15.7	461	0.60	12ZfixCo	Minimal effect compared to 12ZfixC
FWS-DFW (SON)	-278	566	0.58	12ZfixCo	Loss of additional data vs 12ZfixC leads to excessive growth out of troughs at times
GRK-BSM (JJA)	-198	516	0.60	12ZfixCo	Minimal effect vs 12ZfixC since little missing data
HGX-HOU (DJF)	-308	732	0.22	12ZfixCo	Overall, many more reasonable days, even more than 12ZfixA; AMDAR unreliable at times; more aggressive with less surface data vs 12ZfixC and penalized heavily at times for misses
EWX-SAT (MAM)	15.6	473	0.59	12ZfixCm1	Very similar to 12ZfixCo; slightly worse on single day with missing data filled in
FWS-DFW (SON)	-180	521	0.57	12ZfixCm1	Marginally better in cases with missing data that are filled in versus 12ZfixCo
GRK-BSM (JJA)	-246	656	0.56	12ZfixCm1	Very similar to 12ZfixCo but slightly worse on a number of days where additional data prevented growth
HGX-HOU (DJF)	-24.9	676	0.34	12ZfixCm1	A number of days have worse alignment in trough vs 12ZfixCo, likely due to more data at surface resulting in flatlining

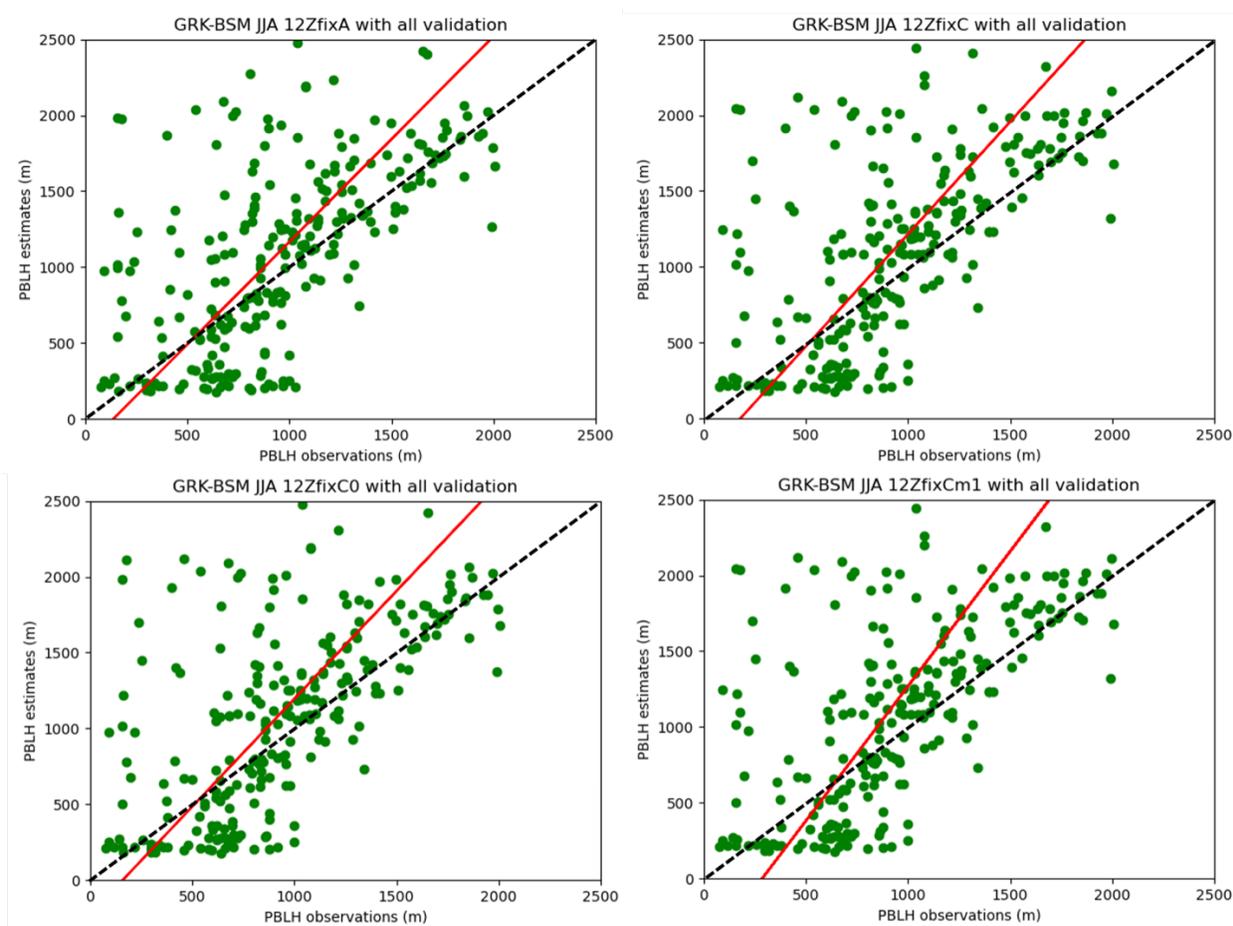


Figure 6. PBLH radar estimates versus AMDAR observations for all four fix experiments using GRK-BSM JJA as an example. Red line is RMA regression fit; black dashed line is 1:1. Radar PBLH algorithm performance across all 12Zfix experiments is consistent when compared to AMDAR PBLH estimates. Clockwise from top: 12ZfixA; 12ZfixC; 12ZfixCm1; and 12ZfixCo.

A combination of competing factors is reflected in the statistics: imperfect AMDAR positions, the dearth of surface data that force higher than optimal heights and successful growth, and also penalties for a small number of days with large errors. The results of the initial experiment, “12ZfixA”, guided the refinements in subsequent experiments. Objective statistical scores and descriptions are presented in Table 8. Note that in Table 8, the Mean Bias Error (MBE) statistics are computed as observed (AMDAR) minus predicted (radar); when the radar algorithm is too low, the bias will be positive. Comments reflect subjective evaluation in the context of the deficiencies stated earlier.

Introduction of 12ZfixA code changes dramatically improves the representation of the PBLH estimates for all site-season cases except for HGX. Generally, the trough is identified much better than for the baseline case, though with a tendency for noisy PBLH traces and some overaggressive growth. For HGX, missing data at the surface impeded substantial progress, but behavior overall was improved. Based on the HGX performance, three additional modifications were introduced (12ZfixC; 12ZfixCo; 12ZfixCm1) which did not appreciably vary from the 12ZfixA results for any site except HGX (Figures 6 and 7).

In all “12Zfix” experiments, the more realistic starting position at 1200 UTC allows the algorithm to find the rising troughs by ignoring the residual layers; meanwhile, the changes to the growth parameters allow for maintenance of the estimates in the *zdr* trough over time. Overall, the “12Zfix” results suggest significant improvement in the reliability of the radar algorithm and thus the utility of the WRF validation code later in the processing stream. Subsequent modifications to the growth parameters (compare 12ZfixA to 12ZfixCo) show modest improvements, with the largest improvement associated with a clear reduction in bias for cases from HGX-HOU, which frequently had less robust radar representation. At this site, the use of only $\frac{1}{4}$ of the full azimuthal range to mitigate influence from the Gulf may play a role in the weaker representation of the trough features. The inclusion of this specific baseline set of days for the HGX radar, and ultimately improving the performance for these marginal cases, is an important consideration.

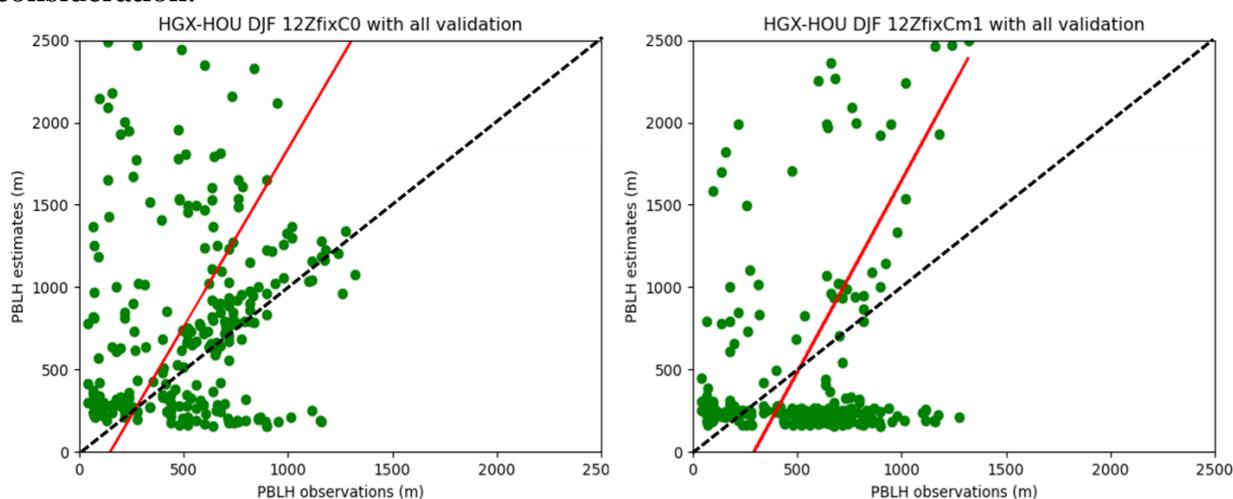


Figure 7. PBLH estimates using (left) 12ZfixCo and (right) 12ZfixCm1 at HGX-HOU for DJF season. Flatlining in 12ZfixCm1 is pronounced. Red line is RMA regression fit; black dashed line is 1:1.

Experiment 12ZfixC additionally changes the filtering threshold for data removal (changes from $zdr < 0$ dB in 12ZfixCo to $zdr < -2$ dB, the value used in Banghoff et al., 2018) and shows the influence of additional surface data. This data may be contaminated by ground clutter from the radar side lobes, but it does not appear unphysical when compared to adjacent pixels and, importantly, allows the radar algorithm to more physically follow near surface values of *zdr*. For the other experiments, the algorithm was forced to position the PBLH estimate on the edge of the missing data and artificially encouraged to find the rising trough in the *zdr* field. One other focus of permitting the low *zdr* values was to fill in the previously empty well-defined *zdr* troughs with data. The impact on the statistics and subjective behavior is minimal except for HGX because the algorithm for 12ZfixA was already still in the vicinity of the centerline of the trough and missing data at the surface is often minimal. For HGX, however, the addition of substantial amounts of previously removed data results in a chronic prevalence of flatlining at the surface. The deeper surface trough inhibits growth, but this results in

better validation scores with the often conservative AMDAR heights, especially when the *zdr* trough for HGX is often poorly defined and the AMDAR heights are unreliable.

The final experiment reported here is 12ZfixCm1 in which the filtering threshold is set at a value of $zdr < -1$ dB, intermediate between experiments 12ZfixCo and 12ZfixC. For this new experiment, the positive impacts of the initial-time height limitation and growth parameters for all sites were preserved while minimizing the bias and improving the correlation of the marginal cases for HGX-HOU. These improved statistics are misleading, however, since they are derived from a dramatic increase in flatlining for this site that compensates for the tendency for the radar height estimates to be higher than the AMDAR observations and leads to a corresponding smaller bias and increase in correlation. The 12ZfixCo experiment (and others) also is penalized for the occasional departure well above the *zdr* trough (while the 12ZfixCm1 experiment has flatlined) and AMDAR locations, which compounds the tendency to often be higher than AMDAR.

In summary, there are dramatic improvements for all 2024 experiments with relatively small differences between the experiments for non-HGX sites. Generally, the “missing data” fix – where the *zdr* threshold is modified to include more surface data – is inconsequential. The 12ZfixCo case was selected as optimal given that (1) it was similarly high performing to all other fix experiments for non-HGX sites; and (2) it provided significant improvements to the special HGX-HOU region. Figure 8 shows a comparison of aggregate performance between Baseline and 12ZfixCo for each site’s seasonal subset. Overall, the 12ZfixCo PBLH estimates are consistently 1.2x-1.5x the AMDAR estimates. Figures 9-12 show specific examples of modifications to PBLH radar trace in the *zdr* field between Baseline and 12ZfixCo cases. With the 12Zfix series, all data prior to 11UTC is irrelevant and ignored. Of note, the dramatic improvement in the single EWX-SAT case (Figure 3) exemplifying pervasive issues in the Baseline algorithm is evident (Figure 9). Ultimately, AER’s goal was to balance overall improvements while not overfitting for a few erroneous cases.

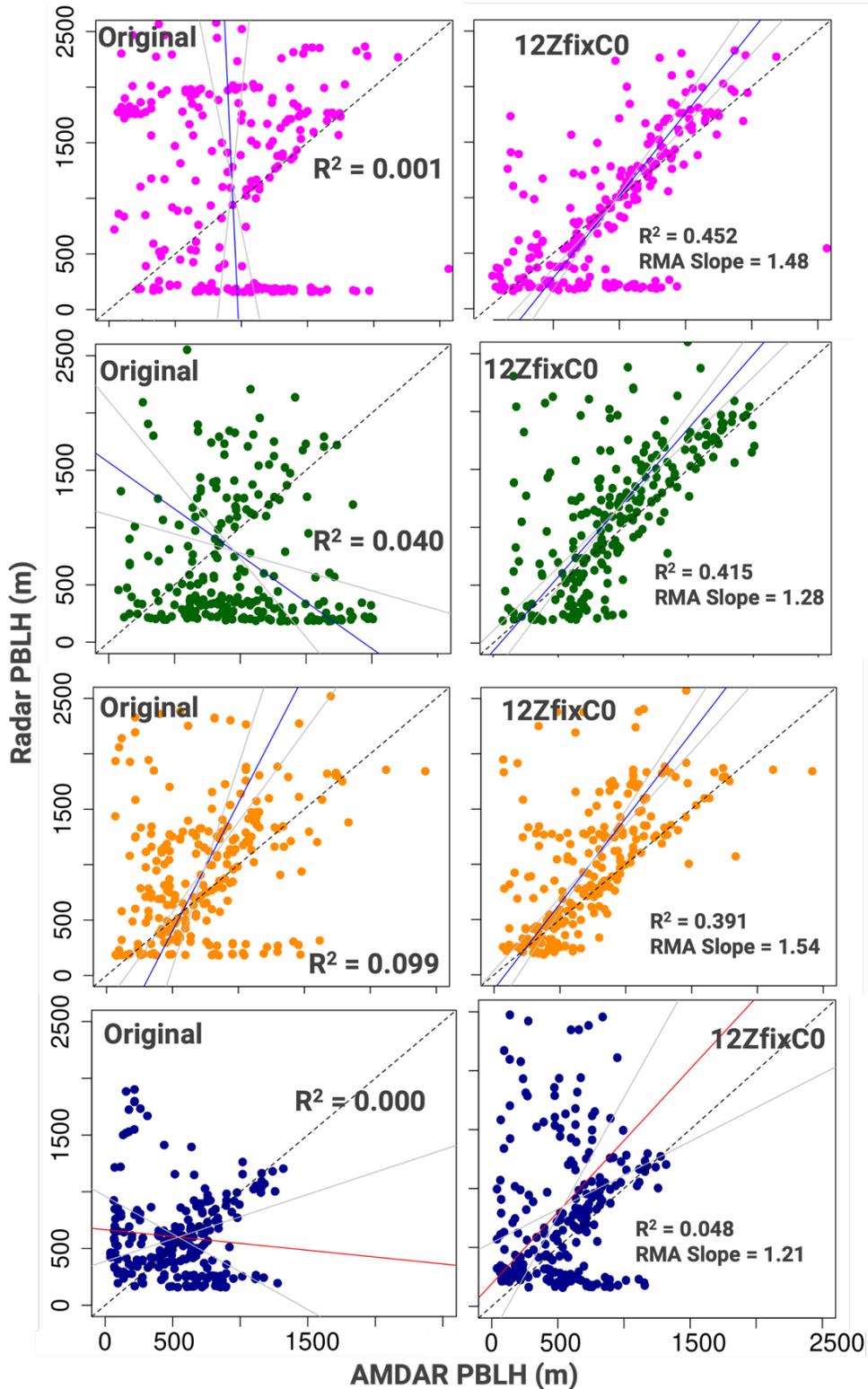


Figure 8. Impact of 12ZfixCo modifications. Scatterplots of radar PBLH estimates vs AMDAR PBLH for each NEXRAD-AMDAR pair’s seasonal evaluation subset. Rows 1-4 exhibit results for EWX-SAT MAM, GRK-BSM JJA, FWS-DFW SON, and HGX-HOU DJF, respectively. Left column shows results using original Baseline code; right column shows 12ZfixCo results.

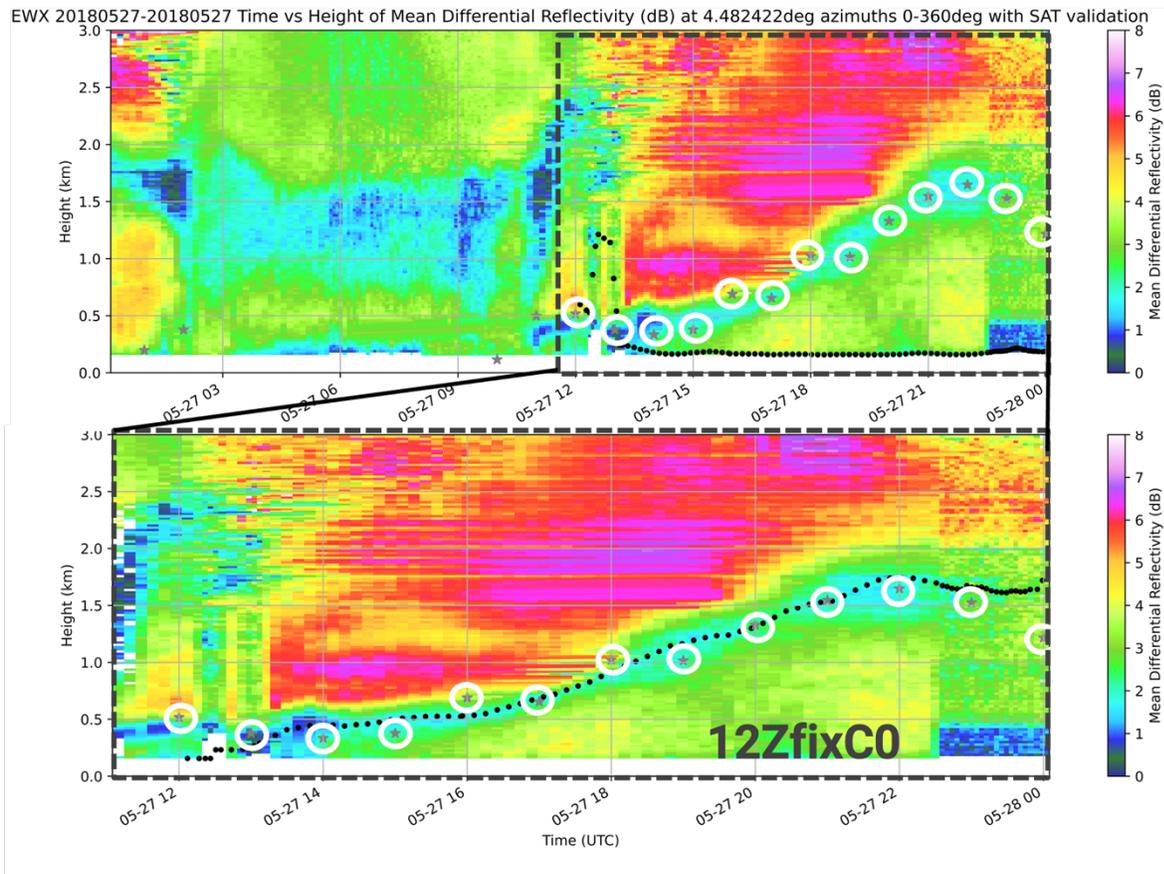


Figure 9. EWX *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from SAT on 27 May 2018. (Top) Reproduction of Baseline case as shown in Figure 3; (Bottom) Correction to flatlining behavior following “12ZfixCo” modifications. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic.

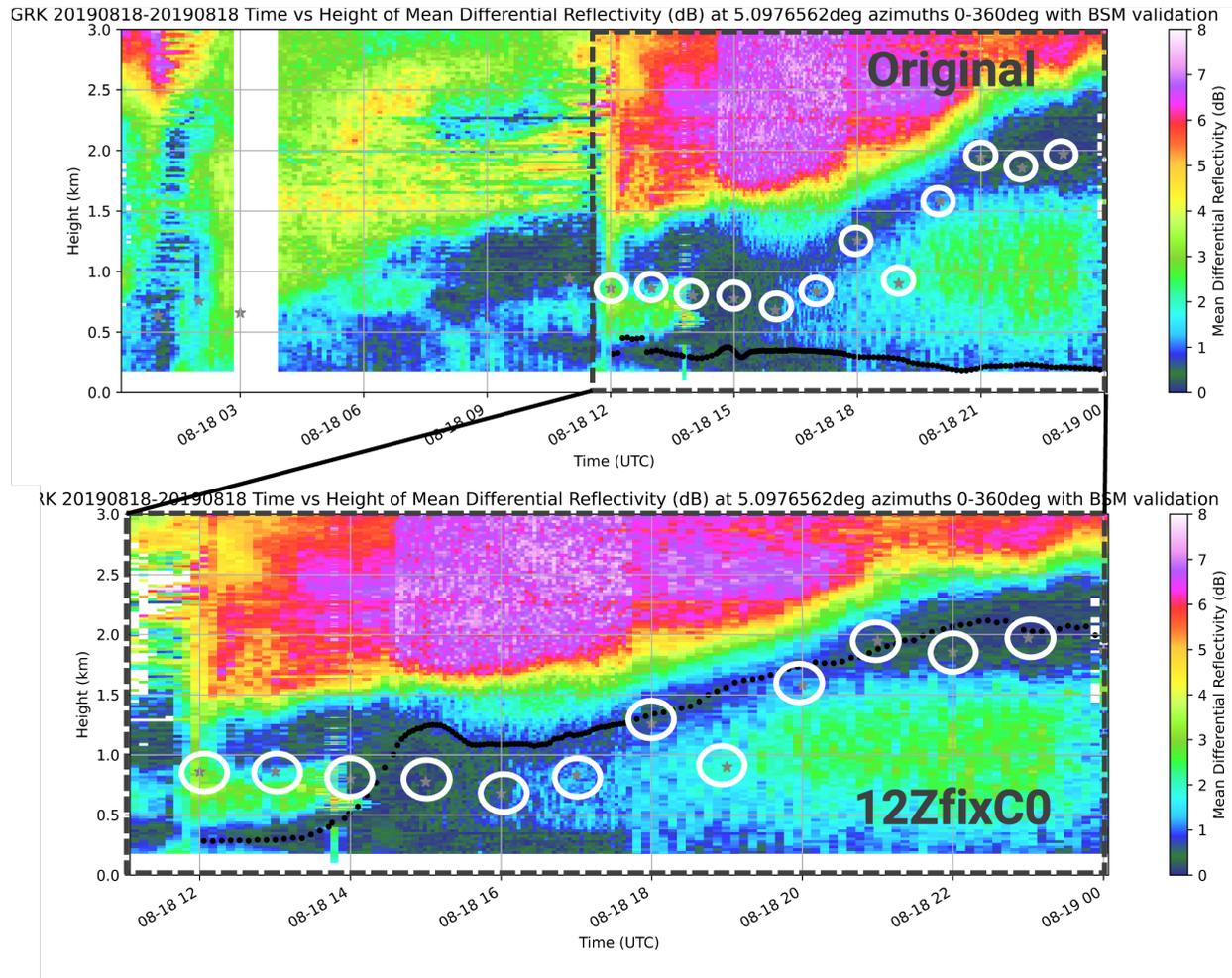


Figure 10. GRK *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from BSM on 18 Aug 2019. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic.

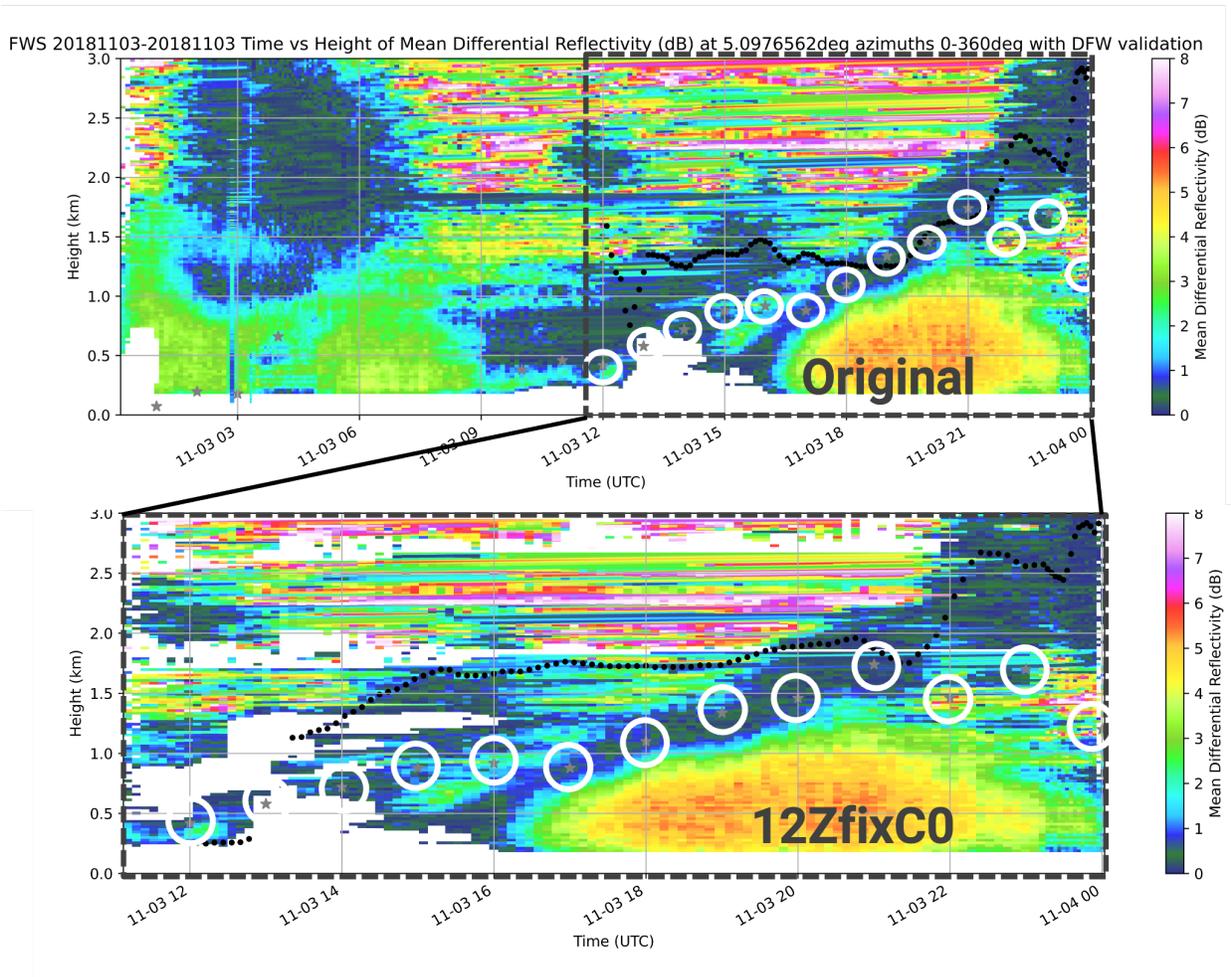


Figure 11. FWS *zdr* field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from DFW on 3 Nov 2018. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic.

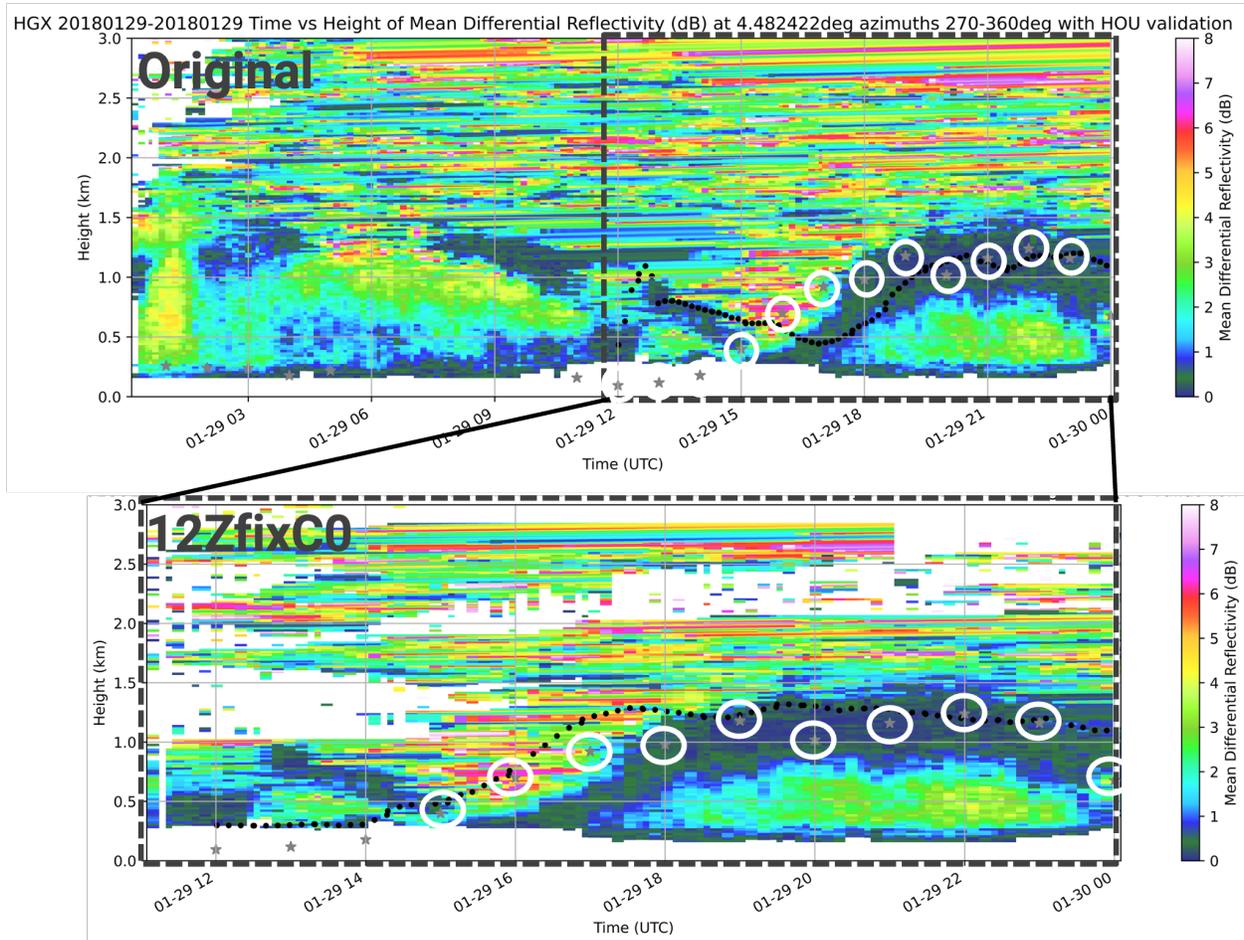


Figure 12. HGX zdr field overlaid with radar PBLH trace (black dots), and AMDAR PBLH estimates (grey stars circled with white for clarity) from HOU on 29 Jan 2018. Note: Baseline graphics include unused data prior to 11UTC; 12Zfix series begins at 11 UTC. Black dashed lines connect the 11UTC to 00UTC (next day) window in the Baseline graphic to the 12Zfix graphic.

3 Validation and Evaluation of the radar-derived PBLH Estimation Methodology in Houston

Task 4 of this work – “*Validation and Evaluation of the radar-derived PBLH Estimation Methodology*” – involved evaluation of the improvements to the radar-derived PBLH algorithm by comparing updated radar-derived PBLH estimates to the suite of Texas TRACER-AQ observations from the 2022 measurement period. TRACER-AQ – TRacking Aerosol Convection interactions ExpeRiment – was an air quality study in Houston, Texas, that explored connections between the area's humid subtropical climate, summer storms, and air quality metrics. It follows that the validation exercise is restricted to the Houston region. However, the use of TRACER-AQ data for radar PBLH algorithm validation poses two challenges: (1) radar algorithm performance in the Houston area is poor relative to other studied regions given the strong Gulf of Mexico influence that requires substantial observation filtering that results in usage of only $\frac{1}{4}$ of the radar azimuthal range; and (2) the TRACER-AQ focus on storm conditions generally

leads to a data set with complex atmospheric conditions that correspond to lower confidence in PBLH.

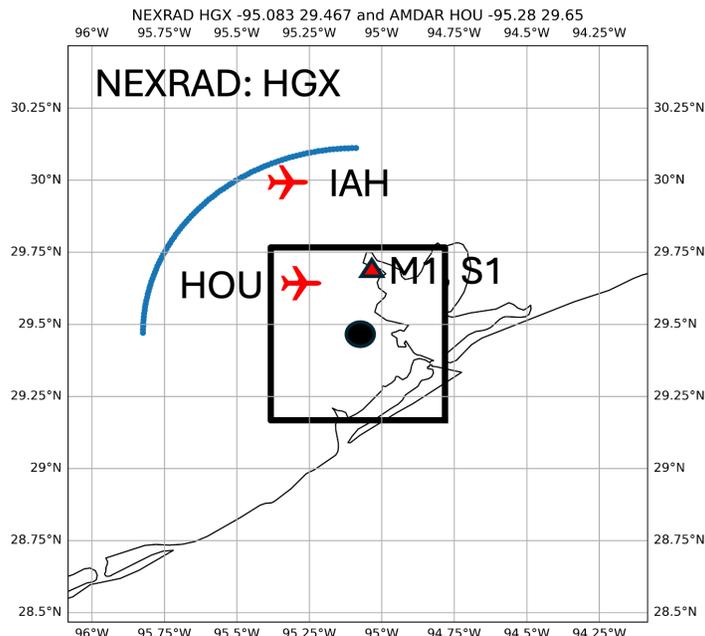


Figure 13. Map of radar and validation sites. Houston NEXRAD (HGX) radar location is indicated by black dot and box surrounding it; ring of locations where radar beam intercepts the PBL top is shown by blue oval and expands or contracts depending on PBLH depth. Red dot shows the AMDAR (airport) locations at HOU (left panel) and IAH (right panel). TRACER-AQ sonde release sites (M1, S1) are at red triangle. Note that for Houston area observations, Gulf influences were filtered and only PBL-radar beam intersections in the northwest quadrant were retained.

3.1 Summary of TRACER-AQ Data

The suite of available TRACER-AQ 2022 data was only available from the Atmospheric Radiation Measurement (ARM) data portal which requires user registration. In addition, the data of interest contained restricted data sets “for testing purposes only” (ARM, 2021c,d) and data sets that have not undergone rigorous quality review (ARM, 2021e a). Table 9 lists all TRACER-AQ PBLH data and their measurement dates for reference and future use; however, for the purposes of the validation exercise in the current work order, only the MCFARL sonde data set (ARM, 2021a,b) was used.

The MCFARL sonde data (ARM, 2021a,b) was identified as the most comprehensive source of external TRACER-AQ 2022 validation data. For each day between 20220101-20220930, there are typically 3-5 sonde releases, each with 4 PBLH estimates from different algorithms: Bulk Richardson 25 pt (BR_25pt); Bulk Richardson 5 pt (BR_5pt); Heffter; and Liu Liang. The data set’s QC flags identify a number of low-quality sondes (QC flag > 0) and these have been excluded from use. The entire MCFARL sonde dataset has been reformatted for use by the time-vs-height radar code. Case studies for the Houston radar (HGX) – the radar closest to the sonde release location - were selected by running the Baseline and 12ZfixCo optimally modified radar algorithm (Henderson, 2024) for the entire dataset in 2022. Ultimately, 39 dates were identified for further exploration: 35 with poor Baseline algorithm performance and four with Baseline algorithm high quality (HQ) performance (Table 10).

Table 9. Summary of all TRACER-AQ 2022 PBLH Data. TRACER-AQ Main Site (M1) refers to location at LaPorte, TX (29.67N, -95.059E); Supplementary Site (S1) is located at M1; Supplementary Site (S3) refers to location at Damon, TX (29.328N, -95.741E). Note that sites M1 and S1 are co-located and for the purposes of this study are treated as identical. ***Only data set used in validation exercise.*

Data Set (Source)	Description	Dates Available
PBLHTSONDE1MCFARL** (ARM 2021a,b)	Planetary Boundary Layer Height Value Added Product: Radiosonde Retrievals for M1 and S3	Daily Files for M1: 2022-01-01 to 2022-10-01 S1: 2022-06-30 to 2022-09-29
PBLHTSONDEYR1MCFARL (ARM 2021c,d) <i>NOT USED</i>	Planetary Boundary Layer Height: Radiosonde Retrievals with yearly output	Files for: M1: 2022-09-18 S1: 2022-08-28
CEILPBLHT (ARM 2021e, 2022a) <i>NOT USED</i>	Ceilometer Planetary Boundary Layer Heights	Daily files for M1: 20220101-20221001; S3: 20220630-20220929.
TBSMERGED (ARM 2022b) <i>NOT USED</i>	Tethered Balloon System merged ceilometer data	Generally 4-5 times daily daytime PBLH estimates in separate files (one or two days are missing) for: S3: 20220603-20220614; 20220702-20220714; 20220802-20220814; 20220902-20220914

Table 10. 39 dates used in TRACER-AQ comparison. Bolded dates showed HQ baseline algorithm performance against TRACER-AQ sonde data.

20220114	20220319	20220327	20220417	20220516	20220605
20220208	20220320	20220329	20220418	20220517	20220611
20220209	20220322	20220403	20220421	20220519	20220619
20220214	20220323	20220407	20220426	20220525	20220917
20220302	20220324	20220408	20220427	20220526	
20220315	20220325	20220409	20220512	20220527	
20220318	20220326	20220414	20220513	20220603	

3.2 Results

Validation statistics comparing the radar PBLH estimates with the four sonde estimates (hereafter, “HGX-MCF”) for the 35 baseline case study days and 4 HQ days (Table 10), separately, have been generated.

Tables 11 and 12 summarize the results of the validation with TRACER-AQ sonde data for case study and HQ days, respectively. The Baseline science algorithm performance is poor for the larger set of days for which the *zdr* trough is present (mean bias range of -769 m to -136 m, with an average of -369 m) and considerably better for the substantially smaller set of excellent days (mean bias up to 167 m, with an average of 39.5 m). Outside

of the HQ days, the baseline algorithm tends to underestimate PBLH relative to TRACER-AQ sonde estimates. Implementing the 12ZfixCo corrections improves the performance across the previously problematic 35 case study days (mean bias range of -199 m to 435 m, with an average of 201 m). While the 12ZfixCo leads to poorer performance on the HQ days (mean bias range 161 m to 375 m; with an average of 247 m), the performance converges toward the 35 case study days suggesting that the HQ days were possibly being overfit by the baseline algorithm at the expense of the bulk performance. Furthermore, the root mean square errors (RMSE) for both versions of the radar algorithm on the HQ days are comparable.

Table 11. Comparison of Radar PBLH estimates with TRACER-AQ PBLHTSONDE1MCFARL dataset for case study (non HQ) days. Comparison with sonde estimates conducted before and after 12ZfixCo (as FixCo) modifications. Comparisons provided for each sonde PBLH method and as average aggregating all sonde estimates. Mean Bias Error (MBE, in meters); Root Mean Square Error (RMSE, in meters) and correlation coefficient (r) summarize base vs FixCo modifications.

	BR_25	BR_25	BR_5	BR_5	Heffter	Heffter	Liu Liang	Liu Liang	Avg	Avg
<i>Version</i>	<i>Base</i>	<i>FixCo</i>	<i>Base</i>	<i>FixCo</i>	<i>Base</i>	<i>FixCo</i>	<i>Base</i>	<i>FixCo</i>	<i>Base</i>	<i>FixCo</i>
MBE (m)	-135.7	434.7	-352.8	225.9	-769.1	-199.3	-262.2	308.3	-369	200.6
RMSE (m)	1052	1246	1136	1240	1447	1260	1055	1278	1096	1176
r	-0.13	0.04	-0.15	0.10	-0.17	-0.15	-0.07	0.0	-0.15	0.03

Table 12. Comparison of Radar PBLH estimates with TRACER-AQ PBLHTSONDE1MCFARL dataset for 4 HQ days only. Comparison with sonde estimates conducted before and after 12ZfixCo (as FixCo) modifications. Comparisons provided for each sonde PBLH method and as average aggregating all sonde estimates. Mean Bias Error (MBE, in meters); Root Mean Square Error (RMSE, in meters) summarize base vs FixCo modifications. Correlation coefficient (r) is not provided due to low sample size (n=4).

	BR_25	BR_25	BR_5	BR_5	Heffter	Heffter	Liu Liang	Liu Liang	Avg	Avg
<i>Version</i>	<i>Base</i>	<i>FixCo</i>								
MBE (m)	38.3	246	-1.69	206	-45.9	161	167	375	39.5	247
RMSE (m)	549	622	543	599	500	574	548	671	520	604

It is important to note the uncertainty among the four sonde PBLH methods themselves: their estimates can vary substantially with differences from 40 m to approximately 400 m (Figure 14). The 12ZfixCo estimates tend to fall within this window, with a mean bias of 200 m. Figure 15 provides an example of the 12ZfixCo algorithm PBLH trace overplotted on the mean *zdr* field for Houston HGX on 20220302 (elevation angle 5.1 degrees, azimuthal range 270-360 degrees). Four TRACER-AQ sondes were obtained on that day, and the variance among the sondes is evident (up to 1000 m).

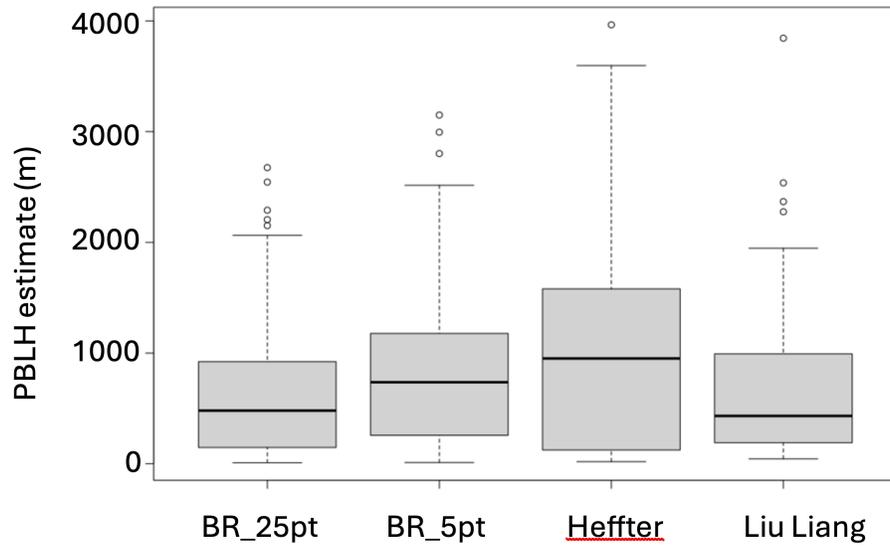


Figure 14. Boxplots of TRACER-AQ sonde-derived PBLH estimates using four methods on the 35 case study days. Mean difference among methods ranges from approximately 40 to 400m.

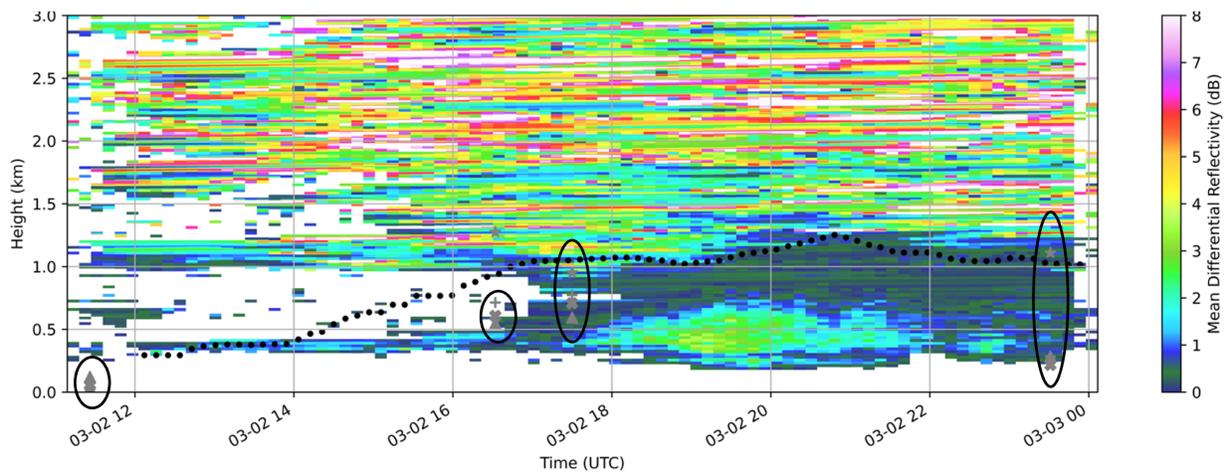


Figure 15. Mean zdr field for Houston HGX on 20220302 (elevation angle 5.1 degrees, azimuthal range 270-360 degrees). 12ZfixCo PBLH trace is indicated by black dots. TRACER-AQ PBLHSONDE1MCFARL estimates are provided by grey shapes, with each shape representing one of four estimation methods. The spread of sonde data estimates is indicated by the black circles.

Overall, implementation of the 12ZfixCo correction to the radar algorithm leads to a substantial improvement in mean bias relative to TRACER-AQ sonde-derived PBLH estimates. The 12ZfixCo has a mean bias of 200 m relative to the TRACER-AQ sondes, which is comparable to differences among the four sonde-derived estimates themselves (40 to 400 m). In contrast, the baseline algorithm has a mean bias of -369 m relative to the sonde estimates.

It is emphasized that the Houston radar, with its Gulf influence and associated restrictive 270-360 degree azimuthal data range, tends to have poor representation of

PBLH. Despite this challenge, the 12ZfixCo correction provides value to Houston area PBLH estimates. Based on comparisons across other Texas cities with more robust radar data representation (discussed further in the Final Report of this Work Order), the 12ZfixCo corrections provide a generalizable and dramatic improvement relative to the Henderson & Vagasky (2023) Baseline algorithm.

4 Quality Assurance

The processing and analysis scripts used in this project were inspected by a team member not involved in their creation for accuracy. All automated calculations and at least 10% of manual calculations were inspected for correctness. This meets the requirement of Level III QAPPs that 10% of the data must be inspected.

As the quality of the information, including secondary data was not evaluated by EPA, the below disclaimer applies to all project deliverables:

Disclaimer: The information contained in this report or deliverable has not been evaluated by EPA for this specific application.

5 Conclusions

Below is a summary of project conclusions, with reference to the corresponding report section.

- Radar algorithm code modifications included addressing improper weighting of *zdr* minima within previous day's (residual) layer; limiting height range to search for *zdr* minima; and a more liberal approach to *zdr* data inclusion. (Section 2).
- The modifications led to dramatic improvements in PBLH estimation when compared to AMDAR data for all three non-Houston sites. (Section 2).
- Of the four versions of modified code, the choice of specific version only mattered for the Houston area where there was 75% less radar representation due to Gulf of Mexico marine boundary layer influence. (Section 2).
- The optimal radar algorithm version across all sites was "12ZfixCo" which typically estimated 1.2-1.5x AMDAR PBLH (Section 2).
- The relaxation of *zdr* data thresholds to address missing data was the only inconsequential modification. (Section 2).
- When compared to TRACER-AQ sonde-based estimates from Houston in 2022, the Baseline algorithm significantly underestimates PBLH. (Section 3).
- While the 12ZfixCo version tends to overestimate PBLH compared to TRACER-AQ 2022 sondes, the mean bias is reduced from the Baseline version. The magnitude of mean bias depends on the sonde PBLH estimation method, which themselves can vary over a 600m range. (Section 3).
- Despite overall improvement (manifested as a reduction in Mean Bias), correlation is poor, likely due to low data availability from the HGX radar. (Section 3).

6 Recommendations for Future Work

Based on the results of this work, the following recommendations are made for further study:

- In-depth analysis of seasonality in radar representation for each site will be helpful for future analyses.
- Ayazpour et al. (2023) recently developed daily PBLH estimates (13:00-14:00 local solar time) over the contiguous US from 2005-2019 which can be used in broader evaluations of the PBLH radar methodology.
- As sonde estimates in other Texas cities becomes available, future work should evaluate performance outside of the Houston area.

7 References

- Atmospheric Radiation Measurement (ARM) user facility. 2021a. Planetary Boundary Layer Height (PBLHTSONDE1MCFARL). 2021-12-31 to 2022-09-19, ARM Mobile Facility (HOU) Houston, TX; AMF1 (main site for TRACER) (M1). Compiled by D. Zhang and D. Zhang. ARM Data Center. Data set accessed 2024-02-20 at <http://dx.doi.org/10.5439/1991783>.
- Atmospheric Radiation Measurement (ARM) user facility. 2021b. Planetary Boundary Layer Height (PBLHTSONDE1MCFARL). 2022-03-02 to 2022-06-21, ARM Mobile Facility (HOU) Houston, TX; Supplemental Facility 1 for instruments colocated at M1 (S1). Compiled by D. Zhang and D. Zhang. ARM Data Center. Data set accessed 2024-02-20 at <http://dx.doi.org/10.5439/1991783>.
- Atmospheric Radiation Measurement (ARM) user facility. 2021c. Planetary Boundary Layer Height (PBLHTSONDEYR1MCFARL). 2021-09-18 to 2022-09-18, ARM Mobile Facility (HOU) Houston, TX; AMF1 (main site for TRACER) (M1). Compiled by D. Zhang. ARM Data Center. Data set accessed 2024-02-21 at <http://dx.doi.org/dx.doi.org/10.5439/1991782>
- Atmospheric Radiation Measurement (ARM) user facility. 2021d. Planetary Boundary Layer Height (PBLHTSONDEYR1MCFARL). 2021-08-28 to 2022-08-27, ARM Mobile Facility (HOU) Houston, TX; Supplemental Facility 1 for instruments colocated at M1 (S1). Compiled by D. Zhang. ARM Data Center. Data set accessed 2024-02-21 at <http://dx.doi.org/10.5439/1991782>
- Atmospheric Radiation Measurement (ARM) user facility. 2021e. Boundary-layer height data with CEIL (CEILPBLHT). 2022-01-01 to 2022-03-04, ARM Mobile Facility (HOU) Houston, TX; AMF1 (main site for TRACER) (M1). Compiled by V. Morris, B. Ermold, D. Zhang and Y. Shi. ARM Data Center. Data set accessed 2024-02-20 at <http://dx.doi.org/10.5439/1095593>.
- Atmospheric Radiation Measurement (ARM) user facility. 2022a. Boundary-layer height data with CEIL (CEILPBLHT). 2022-06-30 to 2022-09-07, ARM Mobile Facility (HOU) Houston, TX; Supplemental Facility 3 for ANC: upwind site for summer iop (S3). Compiled by V. Morris, B. Ermold, D. Zhang and Y. Shi. ARM Data Center. Data set accessed 2024-02-20 at <http://dx.doi.org/10.5439/1095593>.
- Ayazpour, Z., Tao, S., Li, D., Scarino, A. J., Kuehn, R. E., and Sun, K. (2023). Estimates of the spatially complete, observational-data-driven planetary boundary layer height over the contiguous United States, *Atmos. Meas. Tech.*, 16, 563–580, <https://doi.org/10.5194/amt-16-563-2023>.
- Banghoff, J. R., Stensrud, D. J., & Kumjian, M. R. (2018). Convective boundary layer depth estimation from S-band dual-polarization radar. *Journal of Atmospheric and Oceanic Technology*, 35(8), 1723-1733.
- Eure, K. C., Mykolajchuk, P.D., Zhang, Y. Stensrud, D.J., Zhang, F., Greybush, S. J. and Kumjian, M. R. (2023). Simultaneous Assimilation of Planetary Boundary Layer Observations from Radar and All-Sky Satellite Observations to Improve Forecasts of Convection Initiation. *Mon. Wea. Rev.*, 151, 795–813, <https://doi.org/10.1175/MWR-D-22-0188.1>.
- Henderson, J., & Mountain, M. (2022). Using National Weather Service (NWS) Radars to Constrain Planetary Boundary Layer Height (PBLH) Simulations. Final Report to

- the TCEQ for Work Order No. 582-22-31690-012 under Contract No. 582-19-90498. 30 June. Available at <https://www.tceq.texas.gov/downloads/air-quality/research/reports/meteorological/5822231690012-22020630-aer-using-nws-radars-to-constrain-pblh-simulations.pdf>
- Henderson, J., & Vagasky, H. (2023). Development and Implementation of a WRF Planetary Boundary Layer Height Operator. Final Report to the TCEQ for Work Order No. 582-23-42428-017 under Contract No. 582-19-90498. 23 June. Available at <https://www.tceq.texas.gov/downloads/air-quality/research/reports/meteorological/5822342428017-20230623-aer-wrf-pbl-height.pdf>
- Henderson, J. (2024a). Planetary Boundary Layer (PBL) Verification with 2022 TRACER-AQ Field Study Data: Development and Implementation of a WRF Planetary Boundary Layer Height Operator and Improvement to radar-based PBLH estimation methodology. Updated Software and User Manual to the TCEQ for Work Order No. 2 under Contract No. 582-23-45974. 30 April.
- Henderson, J. (2024b). Planetary Boundary Layer (PBL) Verification with 2022 TRACER-AQ Field Study Data. Technical Memo to the TCEQ for Work Order No. 2 under Contract No. 582-23-45974. 30 May.
- Li, D. (2020). AMDAR_BL_PBLH_DATASET (v1.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.3934378>
- Zhang, Y., Sun, K., Gao, Z., Pan, Z., Shook, M. A., & Li, D. (2020). Diurnal climatology of planetary boundary layer height over the contiguous United States derived from AMDAR and reanalysis data. *Journal of Geophysical Research: Atmospheres*, 125, e2020JD032803. <https://doi.org/10.1029/2020JD032803>

8 Appendix A: Time vs Height Radar Code Docker Instructions

The Time-vs-height software that translates radar zdr fields to PBLH using the AER algorithm “12ZfixCo” resides on docker hub:

```
testingp1931/time-vs-height-radar-data
```

No science code resides in the image. All science code and auxiliary run scripts are included in tarball delivered to the TCEQ as part of Deliverable 3.1 on 30 April 2024.

The instructions below show how to pull/build docker image and use it to estimate PBLH from radar using the recent fixes described in the body of this Final Report.

Software Setup (uses Docker image):

1. Untar software package delivered to TCEQ as part of Deliverable 3.1 on 30 April 2024.

```
$ tar -xvf software-and-data-delivery-20240429.tar
```
2. Change the permissions of the entire directory

```
$ chmod -R 777 software-and-data-delivery-20240429
```
3. Set up the docker image in one of two ways.
 - 3a. The first is to pull the image from docker hub:

```
$ docker pull testingp1931/time-vs-height-radar-data
```
 - 3b. The second option is to build from the code provided in the software-and-data-delivery-20240429 directory:

```
$ cd pblh-qvp-radar/docker
$ docker build--no-cache -t time-vs-height-radar-data . >&!
Build.log
```
4. Successful pull/build of the docker image should result in the following:

```
$ docker image ls
REPOSITORY          TAG          IMAGE ID          CREATED          SIZE
time-vs-height-radar-data  latest      9988937d5688    1 day ago      3.94 GB
```
5. An example of running the updated software (12ZfixCo) for GRK-BSM site on 20190708 involves the following:

```
$ docker run --rm --name time-vs-height-radar-data -v
`pwd`::/opt/src -v `pwd`:/data:/opt/data time-vs-height-
radar-data ./time-vs-height-radar-data-args-2024-amdar-
12ZfixC0.py -i GRK-BSM -s 20190708 -e 20190708 -azs 0 -aze
360 >&! time-vs-height-radar-data-12ZfixC0-GRK-BSM-
20190708.txt &
```

And this will output a series of graphics, csv files, and netcdf files to:
12ZfixC0-out/GRK-BSM/