TECHNICAL MEMO

Using National Weather Service (NWS) Radars to Constrain Planetary Boundary Layer Height (PBLH) Simulations

Task 8: Final Report

TCEQ Contract No. 582-19-90498
Work Order No. **582-22-31690-012**

Deliverable 8.2
Revision 1.2

Prepared by:

John Henderson and Marikate Mountain
Atmospheric and Environmental Research, Inc. (AER)
131 Hartwell Ave.
Lexington, MA 02421
Correspondence to: jhenders@aer.com

Prepared for:

Erik Gribbin
Texas Commission on Environmental Quality (TCEQ)
Air Quality Division
Building E, Room 342S
Austin, Texas 78711-3087

30 June 2022

Document Change Record

Revision	Revision Date	Remarks
0.1	10 June 2022	Internal Version for Review
1.0	15 June 2022	Draft Report Delivery to TCEQ
1.1	30 June 2022	Internal Version for Review
1.2	30 June 2022	Final Delivery to TCEQ

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Preamble

The purpose of this project is to deliver software to compute boundary layer height (PBLH) estimates from radar data that can later be used in data analysis and meteorological modeling simulations. Using a novel technique involving NWS radar data, AER computed the depth of the PBL and validated against observations furnished by TCEQ. As part of this work, AER also evaluated the premise of the scientific technique and trained staff on how to run the software.

This Final Report is to be delivered to the TCEQ Project Manager electronically (i.e., via file transfer protocol (FTP) or e-mail) in Microsoft Word and PDF formats no later than the deliverable due date shown below. This Final Report provides a comprehensive overview of activities undertaken and any data collected and analyzed. The Final Report highlights major activities and key findings, provides pertinent analysis, describes encountered problems, and associated corrective actions, and/or detailed relevant statistics, including data, parameter, or model completeness, accuracy, and precision.

Deliverable 8.2: Final Report
Deliverable 8.2 Due Date: June 30, 2022

2 Executive Summary

Banghoff et al. (2018; hereafter, BH18) presented a method to use the differential reflectivity field of NWS radars to estimate planetary boundary layer height (PBLH). Differential reflectivity 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 this project, AER developed and installed a proof-of-concept system that applied the quasi-vertical profile (QVP) methodology used by BH18 to NWS 88D weather radars at selected sites in Texas. The system is based on publicly available Python-language software. An in-depth scientific review of the technique, and evaluation of the pre-existing software packages, was undertaken to ensure that the approach is scientifically sound, and that the underlying software is well-maintained. AER created wrapper scripts in Python as part of the end-to-end system and exercised the code on approximately 50 use cases. We refined the scientific processing steps of BH18 through a thorough evaluation process that leveraged the meteorological knowledge of AER scientists. The system was installed on the TCEQ machine named 'eira' using Docker. This provides a convenient method for porting to other machines as needed. A selection of case studies was run on eira to ensure that the software performed as expected.

The radar based PBL heights were validated against readily available observations using validation software developed by AER. An ozonesonde dataset from the Houston area was provided by TCEQ for this purpose. The small number of days in this dataset, and especially the prevalence of non-optimal meteorological conditions on those days, however, prevented the generation of a usable quantitative assessment of the system.

We recommend that a much broader validation be performed that applies the existing validation software to radiosonde PBLH estimates for each radar site of interest in and near Texas. We anticipate that this effort will document similar system performance to that of BH18. The current proof-of-concept system will also benefit from refinements to the algorithm as the user begins to understand its behavior under a broad range of weather conditions and local influences. We envision it can then be used to validate model based PBLH through development of a forward operator to compute PBLH values from model data at similar spatial and temporal scales as those estimates from the radar.

3 Introduction

The technique applied in this project processes NWS radar data into Quasi-Vertical Profiles (QVPs) that are analyzed algorithmically to extract estimates of PBLH. Below we review the basics of radar for the novice reader and follow with an overview of the scientific approach.

3.1 Basics of NWS Radar

We first describe the US operational radar network and the characteristics of the associated datasets. The operational weather radar network in the US consists of approximately 160 WSR-88D (Weather Surveillance Radars – 1988 Doppler) geographically positioned to maximize their ability to monitor the weather and detect hazards over broad areas and population centers. There are currently 13 radar sites in Texas (see Table 1). The radar hardware rotates on a pedestal, increasing the scan angle every whole revolution, as it emits very short pulses of electromagnetic radiation (S-band, 10-cm wavelength), then listens for backscatter – the energy returned when the pulse reflects off (typically) precipitation and airborne particles. Rayleigh scattering is the predominant source of backscattered energy when the target is less than 7 mm in diameter (Mie scattering dominates for larger particles).

The radar beam is 1° wide and normally is aimed 0.5° above the horizon for the first elevation scan. The beam broadens (width defined where the power reduces to ½ of the peak transmitted power at the centerline) with range – expanding to ~300 m at 10 km and ~3600 m at 120 km. After each revolution, the scan is repeated at successively higher elevations to complete a single volume scan. Each volume scan takes between 5 and 10 minutes depending on whether the radar operator has selected precipitation or clear-air mode, respectively. Clear air mode takes longer on individual elevation angles and operates on fewer angles in total by scanning slower and pulsing longer to maximize the quality of the data, since it is attempting to observe radar returns from very light precipitation, insects, and boundaries in the atmosphere.

Table 1: WSR-88D radars located in Texas.

Radar Name	Identifier
Amarillo	KAMA
Brownsville	KBRO
Corpus Christi	KCRP
Dyess AFB	KDYX
El Paso	KEPZ
Central Texas/Fort Hood	KGRK
Fort Worth	KFWS
Houston/Galveston	KHGX
Laughlin AFB	KDFX
Lubbock	KLBB
Midland/Odessa:	KMAF
San Angelo	KSJT
San Antonio	KEWX

The WSR-88D network, established in the 1990s, was the first operational network to exploit the Doppler shift to generate a radial velocity field. The scientific approach of the current project, in turn, takes advantage of an upgrade during the 2010s that added a vertically-polarized pulse to complement the existing horizontally-polarized pulse – hence the term ‘dual-pol(arized) radar’. This approach improves the ability to characterize the type (e.g., rain vs hail) and shape of precipitation. In addition to the three legacy base datasets of reflectivity (Z), radial velocity (V) and spectrum width (SW), three additional products for meteorological purposes became available: differential reflectivity (ZDR), correlation coefficient (CC) and specific differential phase (KDP). These new products have a wide range of uses relevant to meteorology that include estimation of the amount of liquid water to identification of biota, as well as detection of debris lofted by tornadoes. The spatial resolution of the data varies between products and scan angles, but is either 0.5° to 1.0° azimuthally and 250 m in range with a horizontal range of ~300 km. This resolution results in a single elevation scan, single time dataset of approximately 1200 (radially) x 360 (azimuthally) ‘gates’ or ‘pixels’.

A subset of the above radar products is relevant to the proposed work. The varying characteristics of the reflectivity, differential reflectivity, and correlation coefficient fields in the presence of Bragg scattering (see later), precipitation and biota near the top of the convective PBL can be used to determine PBLH and apply a layer of quality control. The most familiar field is reflectivity factor, z, which represents the scaled average power returned to the radar from all targets in the volume [units of mm^6/m^3 but presented on a log scale, where Z[dBz]= $10\log_{10}(z[10^6/m^3]/1[10^6/m^3]$; range ~-30 to 70+ dBZ]. This field is useful for determining the location, intensity, and movement of precipitation since it depends on the diameter (to the sixth power) and concentration of precipitation particles. Differential reflectivity (ZDR; range of -7.9 to 7.9 dB) represents the difference in reflectivity between the horizontally and vertically polarized pulses and is defined as the log of the ratio of the horizontal (legacy signal polarization) to the vertical (dual-pol) pulses. It is used to determine drop shape, with spherical drops having ZDR=0 dB, horizontally pancake-shaped drops having ZDR= ~1 to 5 dB and vertically oriented snow crystals, for instance, having ZDR<0 dB. Biota tend to produce extremely large ZDR values because of their irregular shapes. Correlation coefficient (CC; range of 0-1.05, dimensionless) is a measure of the consistency of the horizontal and vertical returned power and phase with one another for each pulse and so is a measure of the uniformity of the target. Rain has CC values just less than 1, while non-meteorological (biological) targets have lower values (CC<0.95).

The display of radar data typically takes two forms. The most common is the plan-position indicator (PPI) in which the conical (along a single scanning elevation angle) backscatter information for all fields is projected onto a two-dimensional, geolocated x-y plane with the radar antenna at the center. Figure 1 shows an example of differential reflectivity on a PPI. A second form is the range-height indicator (RHI) in which a vertical slice across all scan angles is displayed. This can show detailed information about the vertical structure of precipitation features; however, this approach is not available in real-time due to processing constraints (Rhyzkov et al., 2016).

The approach used for displaying and analyzing QVPs involves a z-t axis, thus representing multiple single-time 'quasi'-vertical profiles. Please see details in the next section.

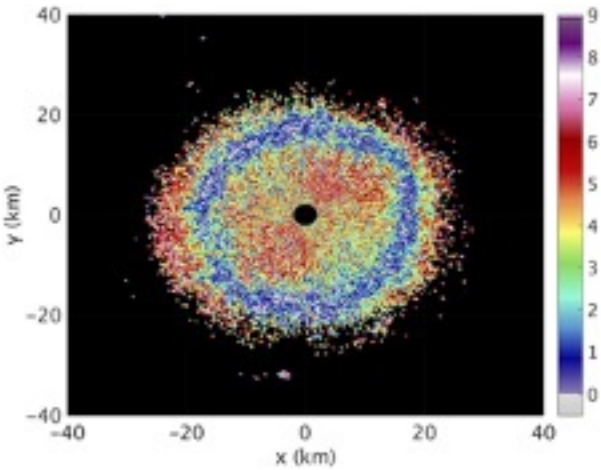


Figure 1: PPI of differential reflectivity (dB; shaded according to scale) at ~4.5 deg elevation at 2345 UTC 11 May 2015 from the KTLX radar. Note clear ring of lower ZDR values (blue) indicative of a Bragg scattering layer. The radar is at the origin. (Reproduction of Fig. 2 from BH18.)

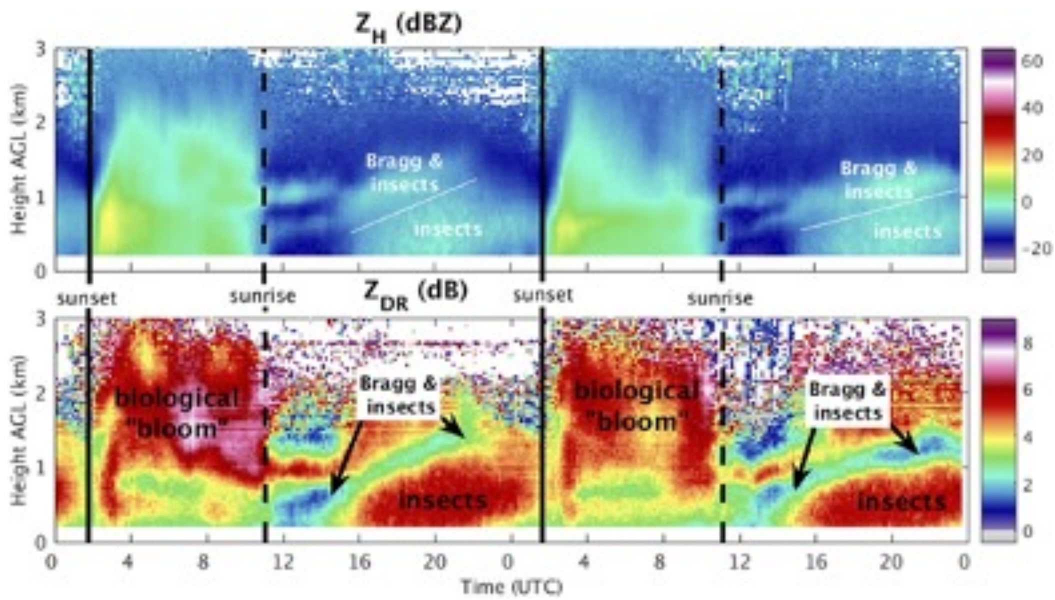


Figure 2: QVPs of Z and ZDR for 20–21 May 2014 at KTLX. Note the daytime evolution of the convective boundary layer (CBL) top characterized by reduced ZDR, and the biota bloom overnight characterized by large Z and ZDR (Reproduction of Fig. 3 from BH18.)

3.2 Scientific Concept

The current project draws heavily from the work of BH18, who applied an analysis and display technique called QVP – quasi-vertical profiles – to the differential reflectivity field (ZDR) generated by the WSR-88D radars. A relative minimum of ZDR owing to Bragg scattering occurs as the radar beam passes from the top of the convective PBL to the free atmosphere under cloud and precipitation-free conditions (see Fig. 1). At the top of the convective PBL, isotropic turbulent mixing of vertical gradients of water vapor mixing ratio and temperature (and corresponding refractive index) can scatter radiation. At scales on the order of 5 cm, Bragg scattering causes local enhancement in the reflectivity field. In clear air, the differential reflectivity and correlation coefficient products can discriminate between these regions of higher reflectivity owing to Bragg scatter and backscatter from non-spherical biota, such as bugs and birds. Pure isotropic turbulent structures that exhibit Bragg scatter have ZDR values near 0 dB and CC near 1.0. In the real atmosphere, this process appears as a layer of lower ZDR while searching vertically from the ground through the top of the PBL. As the PBL deepens in the morning, the height of this layer rises.

3.2.1 Quasi-Vertical Profiles (QVPs) from WSR-88D

The QVP approach in this project involves azimuthal averaging of radar reflectivity (Z), differential reflectivity (ZDR) and cross-correlation coefficient (CC) at (relatively) high antenna elevation to avoid ground clutter, which also increase the angle at which the radar intercepts the top of the PBL. For clear-air scans devoid of precipitation, which are preferred owing to the longer sampling times and increased sensitivity of the radar in this mode, the vertical spacing of gates is ~15 m. The QVPs are formed by combining the single-elevation scans in a height-versus-time format. The azimuthal averaging step reduces the noisiness of the polarimetric variables and assumes that the top of the PBL is homogeneous in all directions. Use of the highest (clear-air) 3.4° to 3.5° scan angles limits the horizontal range from the radar needed for PBL studies to approximately 40 km, assuming a maximum PBL depth of 3 km (~10,000 ft).

The QVP technique applied to ZDR (Figure 2) illustrates the diurnal variation of the Bragg scatter layer. During the development of the convective PBL in the morning, the minimum in ZDR slopes upward to the right with time. A nonuniform PBL top would appear in QVPs as a wider layer of reduced ZDR because of the azimuthal averaging technique employed. Typically, the technique returns two main patterns: 1) a trough of minimum ZDR and 2) a sharp dropoff in ZDR with no higher values atop due to a deep elevated layer of Bragg scatter. In the latter case, definition of the PBLH is more challenging. Linear extrapolation is at times required if the Bragg scatter layer is poorly defined. This approach works well an hour or two before sunset.

3.2.2 High-Level Data Processing Approach

BH18 outlined a nominal approach for computing PBLH:

1) Compute QVP of ZDR field between 1200 and 0000 UTC.
2) Apply quality control procedure in next section.
3) Apply running mean over five time and height steps to reduce noise in time evolution of QVP ZDR during the daytime.

4) Search vertically at each time step at which ZDR minimum occurs; minimum due to Bragg scatter will be biased positively by presence of biotic scattering at each time step.

5) Record the height at which that minimum occurs.

6) Apply running mean to the resulting time series of estimated convective boundary layer (CBL) depth to eliminate large jumps and to produce smoother, and more realistic, behavior.

3.2.3 Quality Control procedure

BH18 then outlined a quality control procedure to address the remaining noisiness of the ZDR QVP by removing individual gates that degrade the ability of the algorithm to determine a physical-looking PBLH estimate:

1) Remove pixels with ZDR < -2 dB owing to ground clutter contamination from sidelobes.
2) Remove pixels with Z > 0 dBZ and CC > 0.8, which indicate precipitation with larger Z values than typical of Bragg scatter. (Note that BH18 tested the sensitivity of the Z threshold by making it 10 dBZ; this led to an insignificant increase of 5 m in the RMSE and no change in correlation in the subsequent analysis.)
3) Remove entire QVP if ZDR < 2 dBZ through the entire column to avoid light precipitation
4) Remove discontinuities defined as a > 1-dBZ difference between adjacent pixels
5) Remove isolated pixels defined as pixels with fewer than two adjacent data points

3.2.4 AER High-Level Data Processing Approach

The following summarizes the high-level approach used by AER to compute PBLH after subjective evaluation of the system in real use cases. Step 4 from BH18 requires several additional refinements:

1) Compute QVP of ZDR field between 1200 and 0000 UTC.
2) Apply quality control procedure in next section.
3) Apply running mean over five time and height steps to reduce noise in time evolution of QVP ZDR during the daytime.

4) Search vertically at each time step for ZDR minimum; minimum due to Bragg scatter will be biased positively by presence of biotic scattering at each time step. Additional requirements added by AER include:

a) Implement a minimum PBLH setting, which screens out QVP values included in the PBLH analysis below a certain height. This option can be tuned by the user. The minimum height requirement helps to avoid ground clutter that caused flat near-surface PBLH estimates.
b) Limit PBLH estimates at 12Z to a maximum of 2000 m to avoid high PBLH values at early times.

c) Implement a maximum growth factor setting, to be tuned by the user, to limit large increases or decreases in PBLH between adjacent times. This also allows time gaps of up to 1 hour between "adjacent" times to avoid large increases or decreases due to short data gaps.

5) Record the height at which that minimum occurs.

6) Apply running mean to the resulting time series of estimated convective boundary layer (CBL) depth to eliminate large jumps and to produce smoother, and more realistic, behavior.

3.2.5 AER Quality Control procedure

The steps below include refinements to those from BH18:

1) Remove pixels with ZDR < 0 dB owing to ground clutter contamination from sidelobes. This was modified from BH18, which removed pixels with ZDR < -2dB, in an effort to remove more ground clutter, which was causing flat near surface PBLH estimates.
2) Remove pixels with Z > 0 dBZ and CC > 0.8, which indicate precipitation with larger Z values than typical of Bragg scatter.
3) Remove entire QVP if ZDR < 2 dBZ through the entire column to avoid light precipitation.
4) Remove discontinuities defined as a > 1-dBZ difference between adjacent pixels.
5) Remove isolated pixels defined as pixels with fewer than two adjacent data points.

4 Project Output Data

The project software generates a) estimates of PBLH in both text and graphics form, plus b) summary statistics when external height observations are available. Here we describe the content and file formats for PBLH estimates and the validation statistics.

4.1 PBLH Estimates

4.1.1 Times Series of PBLH Estimates in CSV Files

The main product deliverable of the project is a CSV-format time series of PBLH estimates. External PBLH validation data are also included if provided at run time via a supplemental file. The CSV file contains estimates for the entire daytime period(s) of radar data specified at run time at time levels corresponding to the valid time of each radar scan. The PBLH product is only computed between 1200 UTC and 0000 UTC. The likelihood of the presence (replacing the missing value) of a PBLH estimate is determined by the variety of quality control and smoothing approaches of the algorithm itself. Typically, an hour-long period during the daytime with a well-formed convective PBL devoid of contamination from cloud layers and precipitation will be sufficient to produce a smoothly changing PBLH estimate. The value 99999 is assigned for missing data.

The output of the CSV file (reproduced below) is directly ingestible without modification into an optional second validation python script. By definition, the fields are separated by commas and represent, respectively, date/time of radar scan in UTC, the PBLH estimate from the algorithm (meters), external validation (if available from external validation file, meters), confidence assigned to external validation (from external validation file) and external validation site name (from external validation file). Note that only one source of external validation can be applied to the PBLH generation code and output in a CSV file, however, all CSV files in the directory can be aggregated for validation purposes via the validation python script.

```
Date,PBLH_est(m),PBLH_val(m),val_confidence,val_site
2021-10-07T00:02:07,99999,99999,99999,99999
2021-10-07T00:11:57,99999,99999,99999,99999
2021-10-07T00:21:47,99999,99999,99999,99999
...
2021-10-07T11:40:52,99999,99999,99999,99999
2021-10-07T11:50:41,99999,99999,99999,99999
2021-10-07T12:00:30,250.0,99999,99999,99999
2021-10-07T12:10:19,250.0,99999,99999,99999
2021-10-07T12:20:08,250.0,99999,99999,99999
2021-10-07T12:29:57,269.0,99999,99999,99999
2021-10-07T12:39:45,329.0,99999,99999,99999
2021-10-07T12:49:34,389.0,99999,99999,99999
2021-10-07T12:59:23,420.0,99999,99999,99999
2021-10-07T13:09:23,465.0,99999,99999,99999
2021-10-07T13:10:32,475.0,99999,99999,99999
2021-10-07T13:20:22,472.0,99999,99999,99999
...
2021-10-07T17:16:05,250.0,99999,99999,99999
2021-10-07T17:25:54,257.0,99999,99999,99999
2021-10-07T17:27:00,99999,1050.0,Moderate,Galveston_Bay
2021-10-07T17:35:43,257.0,99999,99999,99999
2021-10-07T17:45:33,267.0,99999,99999,99999
```

Numerous lines of the above CSV file have been omitted for space, however, groups of lines showing distinct aspects of the file have been separated by the two ellipses. The three lines above the first ellipsis show that the first radar time is 0002 UTC on 7 October 2021. This is the start of the requested two-day period. The scanning strategy of the radar results in each scan taking about 10 minutes. The second radar time is 0011 UTC. The lines following the first ellipsis illustrate the growing boundary layer starting soon after the first application of the QVP approach at 1200 UTC. An example of a validation entry is shown at 1727 UTC below the second ellipsis.

The CSV output resides in a directory named for the radar identifier - e.g., 'out/HGX' - following this nomenclature:

HGX-20211007-20211008-pblh_estimates-**3.5deg**_azimuths_**0-360_Balveston_Bay**_validation.csv, where **HGX** is the radar identifier provided to the script, **20211007-20211008** are the requested beginning and ending dates (here representing two days), **3.5deg** is the scanning angle of the radar, **0-360** is the user-selected choice of azimuths to be used in the QVP technique and **Galveston_Bay** is the (optional) validation source.

4.1.2 QVP Graphics in PNG Files

A graphical representation is generated along with the CSV file (see Figure 3). This PNG format file affords the ability to inspect the overall reasonableness of the algorithm's performance over time, not only with respect to the background differential reflectivity (ZDR) field, but also to optional external validation. The shaded QVP of the ZDR field is shown from the beginning to the end of the requested time period and includes every radar time level. PBLH estimates from the QVP approach are plotted as solid black circles at every time level from 1200 to 0000 UTC when the algorithm completes successfully. External validation is shown by grey symbols. ZDR is presented since the trough in this field is the primary feature used to determine the top of the PBL, subject to quality control constraints from the Z and CC fields. When the top is well-defined and uniform spatially, the trough is sharp. When the trough is broad, this likely shows the deleterious influence of the Gulf sea breeze in Houston, for example, or cloudiness and precipitation.

The filename follows the same nomenclature as the CSV file:

`HGX-20211007-20211008-zdr-3.5deg-time-vs-height_azimuths_0-360_Galveston_Bay_validation.png` is shown in Figure 3.

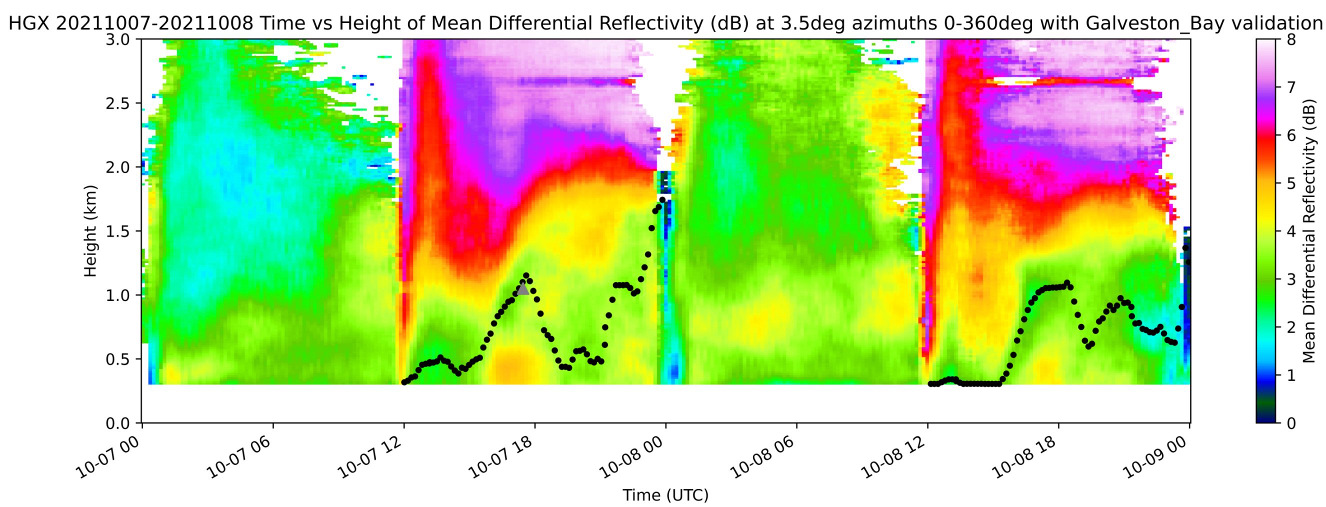


Figure 3 - QVP time-height graphic product for HGX radar for 7-8 October 2021. Example shows two full days of zdr (shaded, dB) with QVP-based PBLH estimates (black circles) applied during 1200-000 UTC periods. Validation is overlaid from Galveston Bay ozonesonde (grey triangle) around 1800 UTC 7 October 2021.

4.2 Validation of PBLH Estimates

4.2.1 Summary Statistics in TXT File

A statistical summary file is output by the validation code. It takes the name:

HGX_with_**all**_validation.txt,

where **HGX** reflects the radar identifier and **all** indicates that all CSV files in the out/ directory were utilized. See details later about the validation code for options. The format of the file (values with dimensions are in meters) is as follows:

```
mbe = 290.0
rmse = 446.88

linear regression:
intercept = 1636.67
intercept_stderr = 0.0

pearson correlation coefficient:
r = -1.0
p-value = 1.0
```

Here, mbe=mean bias error=mean of obs-fcst errors; rmse=sqrt of mean squared obs-fcst errors; intercept=intercept of the linear regression line; intercept_stderr=standard error of the estimated slope; r=Pearson's correlation coefficient; and p-value=two-tailed p-value for testing non-correlation. At least two validation points are required for the validation code to complete successfully.

4.2.2 Scatterplot and Summary Statistics in PNG File

A scatterplot of each PBLH estimate and validation observation is generated in a file similar in name to the above TXT file:

HGX_with_**all**_validation.png.

A sample plot based on only two validation points is shown in Figure 4. The external source of observations is placed on the x-axis, while the radar-based estimates are placed on the y-axis. The confidence of the observations (if available from the source) is coded by color.

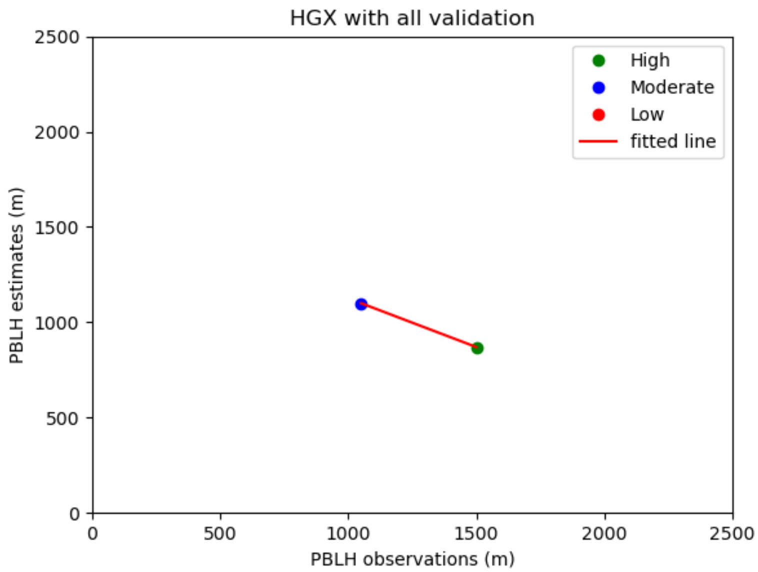


Figure 4 - Scatterplot and best-fit line based on all PBLH estimates and external validation points. Quality of validation points, as determined by the external source, is depicted by color of circle (green=high, blue=moderate, red=low).

5 Software

The project software is intended to be simple to install and use. It has been delivered in a single tarball that contains all required software and also a selection of radar data needed to exercise the software. The underlying software requirements includes only Docker, since all scientific software dependencies for the project deliverables are contained in a single Dockerfile. Additional radar data can be retrieved from the internet as needed. The scientific software underlying the QVP approach is based on publicly available python packages that are retrieved via the Docker build mechanism.

Here we review the pre-existing software and then the modifications and Python scripts developed by the project.

5.1 Pre-Existing Software

5.1.1 Py-ART Software

The Python-ARM Radar Toolkit (Py-ART) software (Helmus et al., 2016; Heistermann et al., 2015; available at arm-doe.github.io/pyart/, documentation at https://arm-doe.github.io/pyart-docs-travis/) ingests WSR-88D radar data and computes QVPs. It is among the list of highlighted utilities for working with WSR-88D data hosted by NOAA NCEI (see list at https://www.ncei.noaa.gov/products/radar/decoding-utilities-examples). The software is open source, comprehensive, well-maintained, documented, and was sufficiently mature and complete to support implementation for the purposes of this project. Use of this software package allowed for the rapid generation of QVPs. Python scripts described below were developed to implement the quality control and smoothing approaches following BH18.

Py-ART is incorporated into the data workflow via a Python module that contains a collection of weather radar algorithms and utilities. Py-ART receives regular updates from the Atmospheric Radiation Measurement (ARM) program, with the latest version (1.11.6) being released on 23 September 2021. The Conda package manager for python currently hosts this latest release for linux, 64-bit Windows and OSX at https://anaconda.org/conda-forge/arm_pyart. Documentation for this latest release is available at https://github.com/ARM-DOE/pyart/releases/tag/v1.11.6. The software has been tested to ensure it is compatible with Python 3.6, 3.7 and 3.8.

5.1.2 Py-ART Software Routines

The following two routines were necessary to process the Level III files available from the NCEI website and the Radar Product Central Collection Dissemination Service (RPCCDS). See details of the data sources below.

A) pyart.io.read_nexrad_level3 – Read a NEXRAD Level 3 product

Inputs - filename (*str*) – Filename of NEXRAD Level 3 product file. The files hosted by at the NOAA National Center for Environmental Information (NCEI; http://www.ncdc.noaa.gov) as well as on the NWS WSR-88D Level III Data Collection and Distribution Network (http://www.roc.noaa.gov/ wsr88d/Level_III/Level3Info.asp) have been tested. (Other arguments have been omitted for space.)
Returns - radar (*Radar obj*) – Radar object containing all moments and sweeps/cuts in the volume. Gates not collected are masked in the field data.

B) pyart.retrieve.quasi_vertical_profile – Creates a QVP object containing fields from a radar object that can be used to plot and produce the quasi-vertical profile

Inputs - radar (*Radar obj*) – Radar object used. (Other arguments have been omitted for space.)
Returns - **qvp** (*Dictionary*) – A quasi-vertical profile object containing fields from a radar object.

5.2 Software Developed for this Project

The Python scientific code developed for this project can be run using wrapper scripts. Below we document the steps required to install and run the end-to-end system. Later, we also supply documentation that the software has been installed on TCEQ's machine of choice and functions as expected. The AER-developed software incorporates the pre-existing software for processing and displaying radar data. There are two main components to the system that are run in this order:

5.2.1 Generate Python environment (run once per machine)

Building a Docker image sets up the environment needed to run the python processing scripts in the next step. This first step uses the file named Dockerfile and should only need to be performed once on a given machine. The output is a Docker image – the precise contents and location of which should not be of concern to the casual user.

5.2.2 Execute Python Scripts (run for each case study)

A) time-vs-height-radar-data.py - This main scientific script ingests Level 3 radar data for one or more dates at a single radar site and optional PBLH validation data (in CSV format). Sample datasets have been provided in the software deliverable. The script generates Quasi-Vertical Profiles (QVPs), then outputs PBLH estimates, and also PBLH validation data if provided at run time, in CSV files. It also creates a timeseries plot containing QVP fields and PBLH estimates, plus the optional validation data.

B) pblh_validation.py – Apply this script to existing use cases for which validation data are available. This script ingests the CSV files generated by time-vs-height-radar-data.py and computes validation statistics, which are then output to text and PNG-format validation files. It requires that time-vs-height-radar-data.py first be executed with optional PBLH observations so that validation data are included in the CSV files output by Step A).

5.3 Execution of Project Software

The steps needed to install and execute the project's software are listed next. As described above, the Python environment is first set up in Docker using the Dockerfile. Then, for each case study, Python scripts are executed.

5.3.1 Install Docker

The steps in section 5.3.1 involve building a docker image to generate the Python work environment that contains the underlying Python libraries needed to run the scripts in section 5.3.2. Section 5.3.1 should be completed once per user per computer. This sets up the run-time environment needed to run the Python code for processing QVPs from Level 3 radar files in the next step and then the subsequent validation. This first step should only need to be performed once. The output is a Docker image – the contents of which should not be of concern to the casual user.

1) **Retrieve tarball of code and data from AER server**

The steps below use a tarball that contains the project code and a variety of radar cases. (Please note that for training purposes on eira, a smaller tarball that contains only a single case study, was used.) Retrieve tarball named "TCEQ-software-deliverable-Task6-WO582-22-31690-012-20220531.tar.gz" from mfts.aer.com by clicking on link emailed to Erik Gribbin on 31 May 2022:

Name: TCEQ-software-deliverable-Task6-WO582-22-31690-012-20220531.tar
Size: 1.4 GB
Expires: August 26, 2022
Download: https://mft.aer.com/download?domain=staff&id=5770dd1266544bdda136d9069d5ed677-c8da468124e34aa0bc5526b1adc6fc6d

2) **Expand tarball in user home directory:**

```
[jhenderson@eira /home/jhenderson]> pwd
/home/jhenderson
[jhenderson@eira /home/jhenderson/docker]> tar -zxf TCEQ-software-
deliverable-Task6-WO582-22-31690-012-20220531.tar
```

A directory named 'docker' will be created that contains the Dockerfile, three python files, sample radar data and ozonesonde validation files. Ensure that the permissions on 'docker' and the three python files are '777' via the chmod command. Note that any directory name can be used; 'docker' can be changed to the user's preference.

3) **Build Docker image**

Build the Docker image: This should only need to be performed once. (Use -–no-cache when building after any changes to force building from scratch; otherwise, Docker will attempt to only update changes.)

```
[jhenderson@eira /home/jhenderson/docker]> cd docker
[jhenderson@eira /home/jhenderson/docker]> docker build --no-cache -t
time-vs-height-radar-data . >&! docker-build-20220526.txt
```

Output will be a Docker image named time-vs-height-radar-data:

```
[jhenderson@eira ~/docker]$ docker image ls
REPOSITORY                  TAG    IMAGE ID     CREATED          SIZE
time-vs-height-radar-data latest 1b385219c862 24 hours ago    2.39 GB
```

5.3.2 Run Scientific PBLH Estimation and Validation Scripts in Docker container

The first of two Python scripts (time-vs-height-radar-data.py) is used to generate Quasi-Vertical Profiles (QVPs) from Level 3 radar data (a sample set of files has been provided), and then a CSV file and QVP graphic are output. The second script (pblh_validation.py) collects the CSV files generated by the previous script to compute validation statistics against the user-specified PBLH obs. The limited validation data available during this project consisted of ozonesondes in the Houston area and has been included in the training dataset. Note that the ancillary file named retrieve_qvp.py must reside in the current working directory. It is used by the other two python scripts during their execution.

1) **Generate PBLH Estimate Files in CSV Format**

```
[jhenderson@eira /home/jhenderson/docker]> 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.py >& docker-
run-QVP-HGX-20210909.txt
```

Creates CSV PBLH files with validation included (if requested) and QVP graphics plots in png format in out/<stid_name>, where <stid_name> is the radar site of interest:

```
[jhenderson@eira ~/docker]$ ls -l out/HGX
-rw-r--r--. 1 polkitd ssh_keys    6502 May 27 13:34 HGX-20210909-
20210909-pblh_estimates-3.1deg_azimuths_270-360_Houston_validation.csv
-rw-r--r--. 1 polkitd ssh_keys 1251158 May 27 13:34 HGX-20210909-
20210909-zdr-3.1deg-time-vs-height_azimuths_270-
360_Houston_validation.png
[jhenderson@eira ~/docker]$
```

The script reads in Level 3 radar files that must already exist in dated directories in data/NCEI. For example:

```
[jhenderson@eira ~/docker]$ ll data/NCEI/HGX/20210727
total 18876
drwxr-xr-x.  2 jhenderson jhenderson 36864 May 26 17:16 .
drwxr-xr-x. 51 jhenderson jhenderson  4096 May 13 13:44 ..
-rw-r--r--.  1 jhenderson jhenderson 14530 Apr 13 10:54
KHGX_SDUS24_N3QHGX_202107270006
-rw-r--r--.  1 jhenderson jhenderson 14251 Apr 13 10:49
KHGX_SDUS24_N3QHGX_202107270015
-rw-r--r--.  1 jhenderson jhenderson 13907 Apr 13 10:55
KHGX_SDUS24_N3QHGX_202107270024
-rw-r--r--.  1 jhenderson jhenderson 13402 Apr 13 10:55
KHGX_SDUS24_N3QHGX_202107270033
-rw-r--r--.  1 jhenderson jhenderson 12726 Apr 13 10:48
KHGX_SDUS24_N3QHGX_202107270042
…
```

The format of the CSV files output is provided in the Section Project .

2) **Validate PBLH Estimates with External Data**

```
[jhenderson@eira /home/jhenderson/docker]> docker run --rm --name
time-vs-height-radar-data -v `pwd`:/opt/src -v `pwd`/data:/opt/data
time-vs-height-radar-data ./pblh_validation.py > & ! docker-run-
validation-HGX-20210909.txt &
```

This creates text-format summary files with statistics and scatterplots in out/validation/:

```
[jhenderson@eira ~/docker]$ ll out/validation
-rw-r--r--. 1 polkitd ssh_keys 22947 May 27 13:51
HGX_with_all_validation.png
-rw-r--r--. 1 polkitd ssh_keys   148 May 27 13:51
HGX_with_all_validation.txt
[jhenderson@eira ~/docker]$
```

The validation script identifies all the CSV files produced by time-vs-height-radar-data.py in a given directory (i.e., out/<stid_name>/*csv.). The CSV files are assumed to be organized into directories by radar site. Only the CSV file names that match the "obsstr" string designation (see Section "Options for Pblh_validation.py" below) will be used in a given validation analysis (e.g. all observations, Houston observations). The validation script reads in the CSV files, calculates statistics that are output to a CSV text file, and creates a scatterplot with a trend line that is output to a PNG file. The statistics included in the text file are mean bias error and root mean square error; the intercept and standard error of the estimated slope from a linear regression; and Pearson's correlation coefficient and p-value for testing non-correlation.

3) **Inspect Container (if needed)**

Entering the Docker container at a shell prompt allows the user to execute the python scripts manually within the supportive environment. This can be useful for debugging the processing of radar data, generation of QVPs, plotting and validation.

```
[jhenderson@eira /home/jhenderson/docker]> docker run --rm -it --name
time-vs-height-radar-data -v `pwd`:/opt/src -v `pwd`/data:/opt/data -v
`pwd`/out:/opt/out --entrypoint '/bin/bash' time-vs-height-radar-data
```

A script from section 5.3.2 can then be run at the command line via 'python <script name>'.

5.4 User-Defined Options

These options are controlled by the user and must be set prior to execution of the two Python scripts.

5.4.1 Options for Time-vs-height-radar-data.py

These options exist for generating QVPs and should be adjusted as needed by the user. Default values are provided.

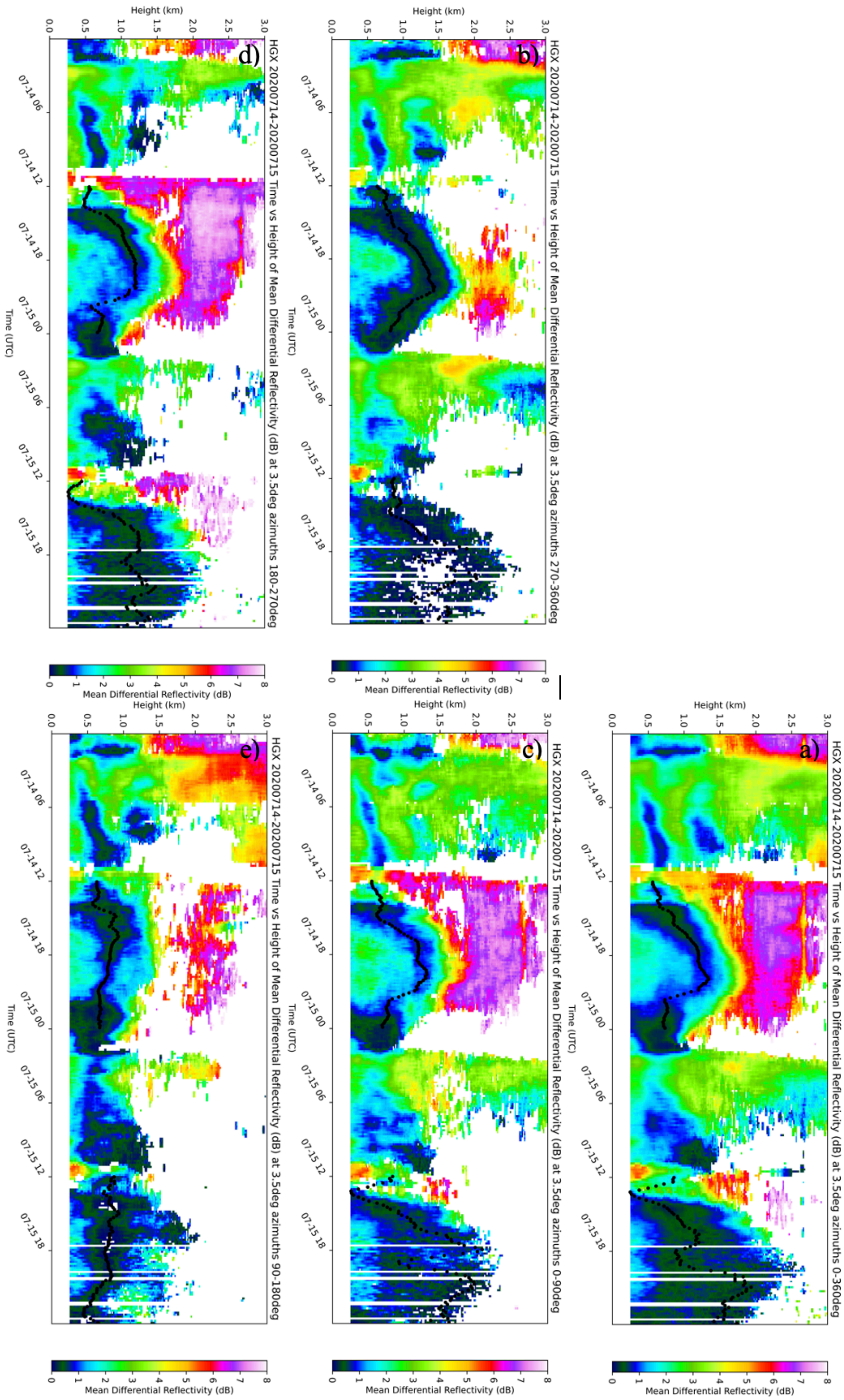
```
#------------------------- USER OPTIONS SECTION ----------------------
# list of date(s) to process and plot
datestr = ['20210907','20210908','20210909','20210910','20210911']
sitestr = 'HGX' #HGX: Houston; EWX: San Antonio; EPZ: El Paso
workdir = '/opt/src/'
```

```
indir = workdir +'data/NCEI/'+sitestr+'/'
outdir = workdir+'out/'+sitestr+'/'

verbose = True
maxtimedelta = 10   #max minutes allowed between radar scans before
data gaps are filled with 99999
maxgrowthfactor = 1.5 #max multiple that pblh estimate is allowed to
grow/shrink in consecutive scans
minheight = 250    # height in meters of lowest data included in
analysis to eliminate ground clutter
scan_directions = [270,360]  #specify azimuth angles [min,max] from
radar scan to include in qvp calculation

# option to plot with validation data
# NOTE: validation_file='none' => validation plotting off
valsite = 'Houston' #Houston, La_Porte, Galveston_Bay, Gulf
#validation_file = 'none'
validation_file = workdir+'data/TCEQ/NASA-ozonesondes/PBLH_estimates-
'+valsite+'.csv'

# file string specifying product and scan angle
filetypes = ['N3X']
#filetypes = ['N3Q','N3C','N3X']
fldsinfile = {'N3Q':['reflectivity',],
'N3C':['cross_correlation_ratio',],
'N3X':['differential_reflectivity',]}
#---------------------------------------------------------------------
```

5.4.2 Sector Analysis in Time-vs-height-radar-data.py

An analysis of the sensitivity to inhomogeneities in the PBL can be performed by restricting the azimuths that are processed by the technique. The default of 0 to 360 degrees can be adjusted via the ‘scan_directions’ option to exclude azimuths. Importantly, azimuths involving permanent obstructions to the radar beam from mountains or pronounced ground clutter (e.g., nearby wind farms) can be excluded. Examples of the sensitivity and resulting improvements in PBLH estimates for cases in Houston (**Error! Reference source not found.**), San Antonio (Figure 6) and El Paso (Figure 7) are shown below.

Figure 5 - (next page) QVP of differential reflectivity and corresponding PBLH estimate for Houston for 14-15 July 2020 computed from azimuthal averages of a) 0-360, b) 270-360, c) 0-90, d) 180-270 and e) 90-180 degrees.


HGX 20200714-20200715 Time vs Height of Mean Differential Reflectivity (dB) at 3.5deg azimuths 0-360deg
HGX 20200714-20200715 Time vs Height of Mean Differential Reflectivity (dB) at 3.5deg azimuths 270-360deg
HGX 20200714-20200715 Time vs Height of Mean Differential Reflectivity (dB) at 3.5deg azimuths 0-90deg
HGX 20200714-20200715 Time vs Height of Mean Differential Reflectivity (dB) at 3.5deg azimuths 180-270deg
HGX 20200714-20200715 Time vs Height of Mean Differential Reflectivity (dB) at 3.5deg azimuths 90-180deg
a)
b)
c)
d)
e)
Height (km)
Time (UTC)
Mean Differential Reflectivity (dB)

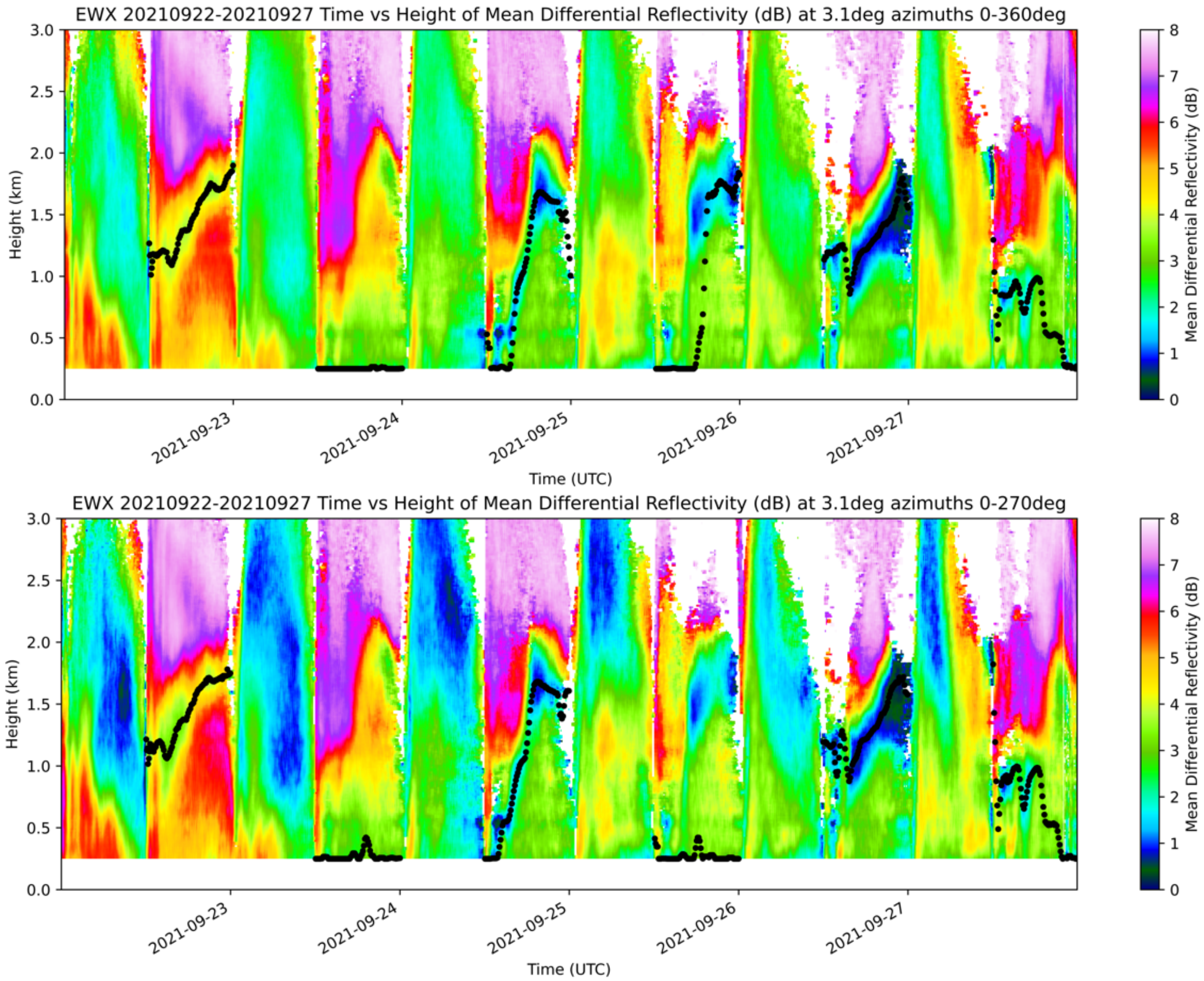


Figure 6 - As in **Error! Reference source not found.**, but for San Antonio from 22-27 September 2021 for azimuthal ranges of (top) 0-360 and (bottom) 0-270 deg.

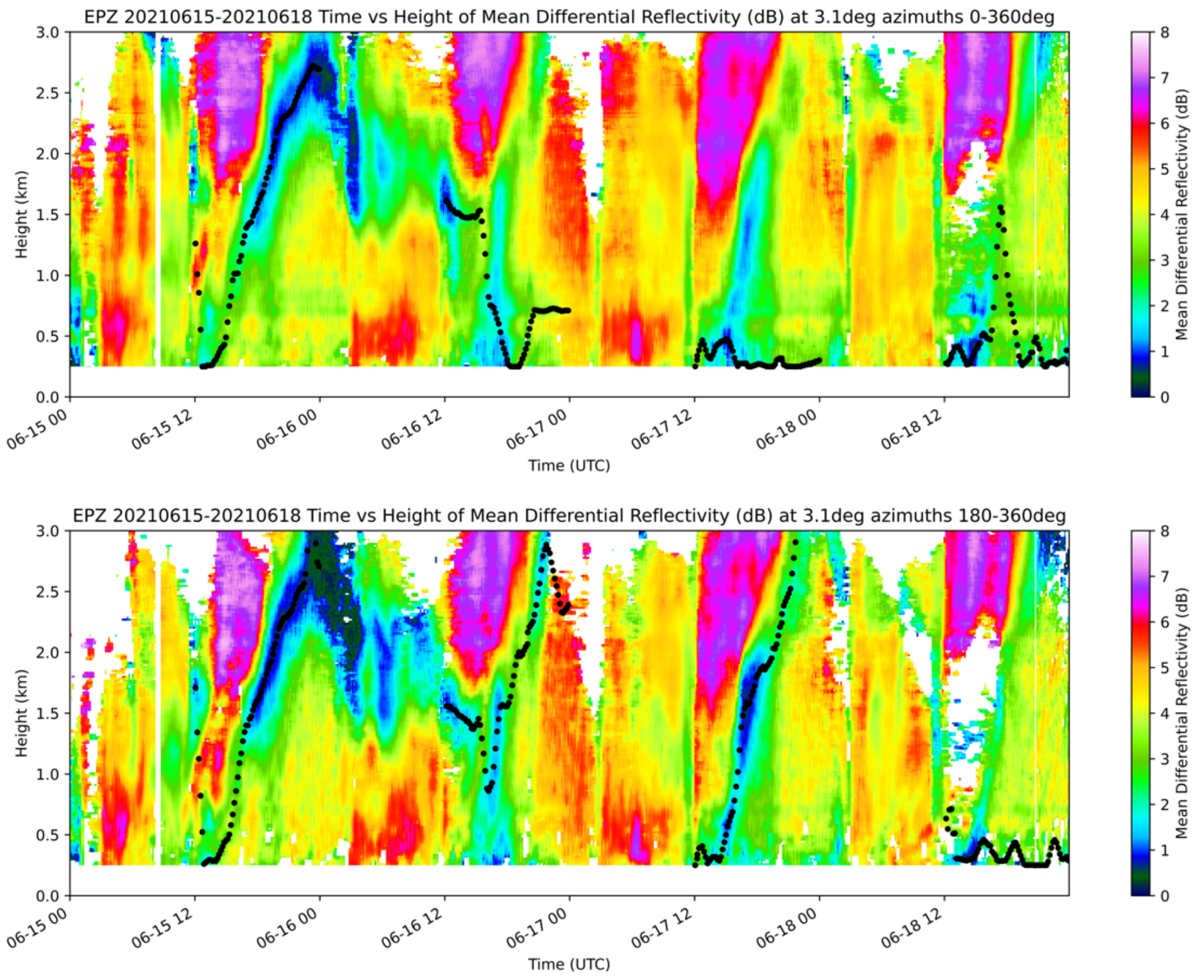


Figure 7 - As in **Error! Reference source not found.**, but for El Paso from 15-18 June 2021 for azimuthal ranges of (top) 0-360 and (bottom) 180-360 deg.

5.4.3 Options for Pblh_validation.py

These options exist while validating the QVP PBLH estimate files and should be adjusted as needed. Defaults are provided.

```
#------------------------- USER OPTIONS SECTION ----------------------
workdir = '/opt/src'
sitestr = 'HGX'  # radar site
sitedir = workdir+'out/'+sitestr+'/'
obsstr = 'all'   # files to use: all, Houston, Galveston_Bay,
La_Porte, Gulf
outdir = workdir+'out/validation/'
maxtimedelta = 60   #max minutes allowed between radar scan and
validation time
ylim = np.array([0,2500])   # default => np.array([np.99999,np.99999])
#----------------------------------------------------------------------
```

5.5 Existing Installation on TCEQ machine

Here we list the directory structure and files present after a successful installation and execution of project software on the TCEQ machine ‘eira’:

```
[jhenderson@eira ~/docker]$ pwd
/home/jhenderson/docker
[jhenderson@eira ~/docker]$ ll
total 264
drwxrwxrwx.  5 jhenderson jhenderson   4096 May 31 13:49 .
drwx------. 17 jhenderson jhenderson   4096 May 31 13:49 ..
drwxrwxr-x.  4 jhenderson jhenderson   4096 May 27 13:01 data
-rw-rw-r--.  1 jhenderson jhenderson  86468 May 26 17:40 docker-build-
20220526.txt
-rw-r--r--.  1 jhenderson jhenderson    772 May 11 15:27 Dockerfile
-rw-rw-r--.  1 jhenderson jhenderson 101853 May 27 14:09 docker-run-
QVP-HGX-20210909.txt
-rw-rw-r--.  1 jhenderson jhenderson    678 May 27 13:51 docker-run-
validation-HGX-20210909.txt
drwxr-xr-x.  4 polkitd    ssh_keys     4096 May 27 13:45 out
-rwxr-xr-x.  1 jhenderson jhenderson   6706 May 27 13:51
pblh_validation.py
drwxr-xr-x.  2 polkitd    ssh_keys     4096 May 27 13:32 __pycache__
-rw-r--r--.  1 jhenderson jhenderson   4486 May 26 15:30
retrieve_qvp.py
-rwxr-xr-x.  1 jhenderson jhenderson  27026 May 27 14:05 time-vs-
height-radar-data.py
[jhenderson@eira ~/docker]$

[jhenderson@eira ~/docker]$ ll */*

data/NCEI:
total 5800
drwxr-xr-x.  7 jhenderson jhenderson    4096 May 26 17:16 .
drwxrwxr-x.  4 jhenderson jhenderson    4096 May 27 13:01 ..
drwxr-xr-x. 28 jhenderson jhenderson    4096 May 26 17:17 EWX
drwxr-xr-x. 51 jhenderson jhenderson    4096 May 13 13:44 HGX
-rw-r--r--.  1 jhenderson jhenderson 5899974 Apr 15 07:50
radar_data_availability.docx
-rw-r--r--.  1 jhenderson jhenderson     853 Feb 16 10:01 README.txt

data/TCEQ:
total 28
drwxr-xr-x. 3 jhenderson jhenderson 4096 May 16 14:11 .
```

```
drwxrwxr-x. 4 jhenderson jhenderson 4096 May 27 13:01 ..
drwxr-xr-x. 2 jhenderson jhenderson 4096 May 26 17:16 NASA-ozonesondes
-rw-r--r--. 1 jhenderson jhenderson  397 Apr 13 10:05 NASA-TRACER-AQ-
LaPorte-Ceilometer-README.txt

out/HGX:
total 1244
drwxrwxrwx. 2 polkitd    ssh_keys      4096 May 27 14:36 .
drwxr-xr-x. 4 polkitd    ssh_keys      4096 May 27 13:45 ..
-rw-rw-r--. 1 jhenderson jhenderson     403 May 27 14:06 docker-run-
QVP-HGX-20210909.txt
-rw-r--r--. 1 polkitd    ssh_keys      6489 May 27 14:09 HGX-20210909-
20210909-pblh_estimates-3.1deg_azimuths_270-360_Houston_validation.csv
-rw-r--r--. 1 polkitd    ssh_keys   1251158 May 27 14:09 HGX-20210909-
20210909-zdr-3.1deg-time-vs-height_azimuths_270-
360_Houston_validation.png

out/validation:
total 36
drwxrwxrwx. 2 polkitd ssh_keys  4096 May 27 13:50 .
drwxr-xr-x. 4 polkitd ssh_keys  4096 May 27 13:45 ..
-rw-r--r--. 1 polkitd ssh_keys 22947 May 27 13:51
HGX_with_all_validation.png
-rw-r--r--. 1 polkitd ssh_keys   148 May 27 13:51
HGX_with_all_validation.txt
[jhenderson@eira ~/docker]$ date
Fri May 27 15:01:58 CDT 2022
[jhenderson@eira ~/docker]$
```

6 Data Sources

6.1 Radar Data Sources

We have identified two data sources that can provide access to the required radar data for all times from the current (real-time) to the beginning of the polarimetric data archive in the 2010s. Build 12 of the radar internal software accommodated the generation of polarimetric products and was implemented nationwide between 2010 and 2013.

6.2 NCEI - https://www.ncdc.noaa.gov/nexradinv

To incorporate this source of archived radar data, the user will request daily data from the website and manually retrieve the files via the methods offered. Data are typically made available within 30 minutes of initiating the web-based request. Required data products include the highest scan angle (typical angles are included here) Level III N3Q (reflectivity, 460-km range, 3.4 degrees), N3X (differential reflectivity, 300-km range, 3.4 degrees) and N3C (correlation coefficient, 300-km range, 3.4 degrees). An entire day of ~10-minute files for differential reflectivity (sample filename: KHGX_SDUS84_N3XHGX_202201172313) requires 1.5 MB of disk space.

```
# To download radar data files for Houston
# 1) Request NCEI level 3 radar data from:
#       https://www.ncdc.noaa.gov/nexradinv/chooseday.jsp?id=khgx
#       Select: Level-III (Products) (ALL) and date
#
# 2) After receiving order ID in email download files (replace <EMAIL>
and <ORDER_ID>)
# wget --ftp-user=anonymous --ftp-password=<EMAIL> -nv -np -nH --cut-
dirs=100 -r -A "*N3X*" -A "*N3Q*" -A "*N3C*"
ftp://ftp.ncei.noaa.gov/pub/has/<ORDER_ID>/
#
#    This command only downloads files of interest:
#    N3X (differential reflectivity, 300-km range, 3.4 degrees)
#    N3Q (reflectivity, 460-km range, 3.4 degrees)
#    N3C (correlation coefficient, 300-km range, 3.4 degrees)
```

This approach was used to obtain all radar datasets that we used in the current work. It is preferred for the radar to be in clear-air mode for the QVP technique to benefit from the enhanced radar sensitivity and lack of precipitation echoes, though the technique will still execute with precip mode files. The mode of operation for the radar can be assessed on the NCEI web site:

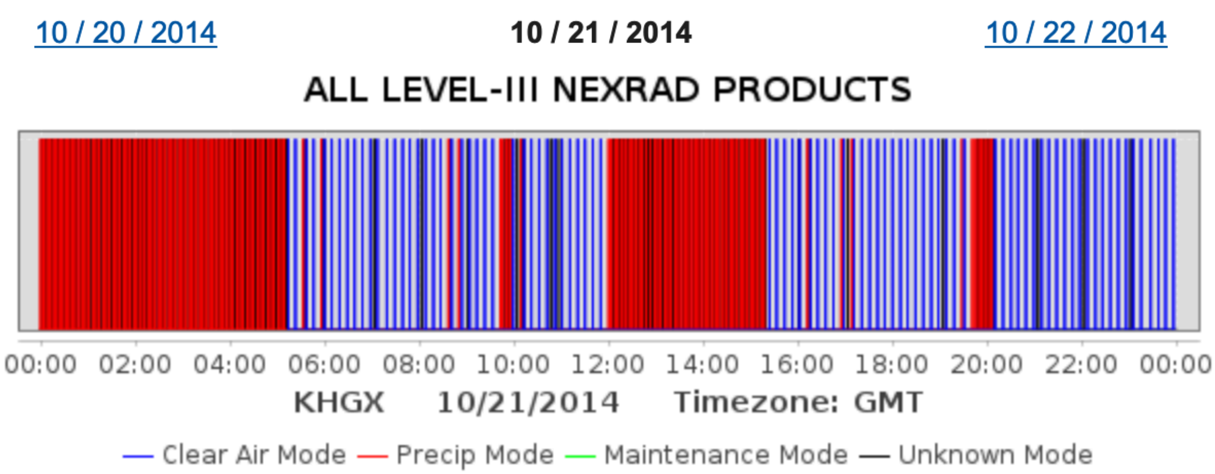


Figure 8 – NCEI graphic indicating the radar scans performed in clear air mode (blue) and precipitation mode (blue).

6.3 NWS WSR-88D Level III Data Collection and Distribution Network

This source of real-time products hosted by the NWS contains 10-s received rotating files in subdirectories identified by product type and radar site: https://tgftp.nws.noaa.gov/SL.us008001/DF.of/DC.radar/DS.p94r3/SI.khgx. For a radar other than Houston (KHGX), replace 'SI.khgx' with the identifier of the site of interest from Table 1. For a product other than differential reflectivity, replace directory path 'DS.p94r3' with the directory containing the product of interest below:

```
DS.p94r3/  Data Subcategory: Base reflectivity - 248 nm Range (angle =
3.1°/3.4°/3.5°)
DS.159x3/  Data Subcategory: Digital Differential Reflectivity (angle
= 3.1°/3.4°/3.5°)
DS.161c3/  Data Subcategory: Digital Correlation Coefficient (angle =
3.1°/3.4°/3.5°)

The files represent a rolling archive over the past 42 h and can be
retrieved via wget:

sn.0214    26-Jan-2022 22:29     4.7K
sn.0215    26-Jan-2022 22:38     4.7K
sn.0216    26-Jan-2022 22:47     4.6K
```

6.4 Validation Data

6.4.1 External PBLH validation data

Sample ozonesonde PBLH estimates provided by TCEQ are included for demonstration purposes. The csv files in data/TCEQ/NASA-ozonesondes are extracted from PBLH_estimates.xlsx and read by time-vs-height-radar-data.py. The data format expected by the python script therefore currently follows that of these worksheets. A different source of validation data will require changes to either the dataset or the ingest code. Validation data are included in the output CSV files in out/<stid_name> that can be used as input by pblh_validation.py.

The format of the file is as follows:

```
[jhenderson@eira ~/docker]$ head data/TCEQ/NASA-
ozonesondes/PBLH_estimates-Galveston_Bay.csv
Flight Number,Date,Launch Time (UTC),Launch Time (CST),PBLH
(km),Confidence,Other Possiblities or notes,
gb001,6/10/2021,15:34,9:34,0.4,Moderate,,
gb002,7/13/2021,15:04,9:04,0.85,High,,
gb003,7/13/2021,20:04,14:04,0.65,High,,
```

```
gb004,7/18/2021,14:54,8:54,0.7,High,,
gb005,7/18/2021,18:09,12:09,0.9,Moderate,,
gb006,7/21/2021,14:52,8:52,0.35,Moderate,Surface layer,
gb007,7/21/2021,18:52,12:52,0.6,Moderate,0.95,1.5
gb008,7/22/2021,14:29,8:29,0.6,Moderate,,
gb009,7/26/2021,14:37,8:37,0.31,High,,
[jhenderson@eira ~/docker]$
```

Note that estimates of the quality of the observation were determined by whomever generated the data (NASA in this case).

7 Validation of PBLH Estimate Product from QVP Technique

Here we report on validation performed by BH18 and then by AER for the project.

7.1 BH18 Validation of PBLH Estimates

BH18 performed a validation exercise over the calendar year 2014 to determine the frequency that a 'usable' day occurred and to compute the error of their PBLH estimates compared to radiosonde datasets. They compared their PBLH estimates based on the 2300 UTC ZDR QVP profiles against the 0000 UTC balloon launches from KOUN (Norman, OK), which is 23 km from the KTLX radar site. The 0000 UTC balloons are released at approximately 2300 UTC and the study focuses on the lower troposphere that is sampled soon after launch. It is expected that validation against 1200 UTC radiosondes would be less reliable in Texas since the convective boundary layer is poorly developed during that part of the morning.

The evaluation approach was applied to each usable day of 2014 based on the absolute maximum vertical gradients for rawinsonde data and the absolute minimum ZDR value for radar data. Manual quality control of both methods led to 243 useful cases (67% of all days) for analysis, with even distribution across seasons (DJF 62, MAM 59, JJA 62 and SON 60 days). There were 122 bad days (33% of all days): 10% (3% of all days) did not have a 0000 UTC sounding; 50% (17% of all days) had a saturated layer from clouds/precipitation near the surface on soundings, indicating that the CBL was indiscernible or absent; 40% (13% of all days) had inconclusive ZDR signatures. Figure 9 is a scatterplot of all PBLH comparisons during 2014 at KTLX on the 243 usable days. The correlation is 0.90 with an RMSE of 254 m.

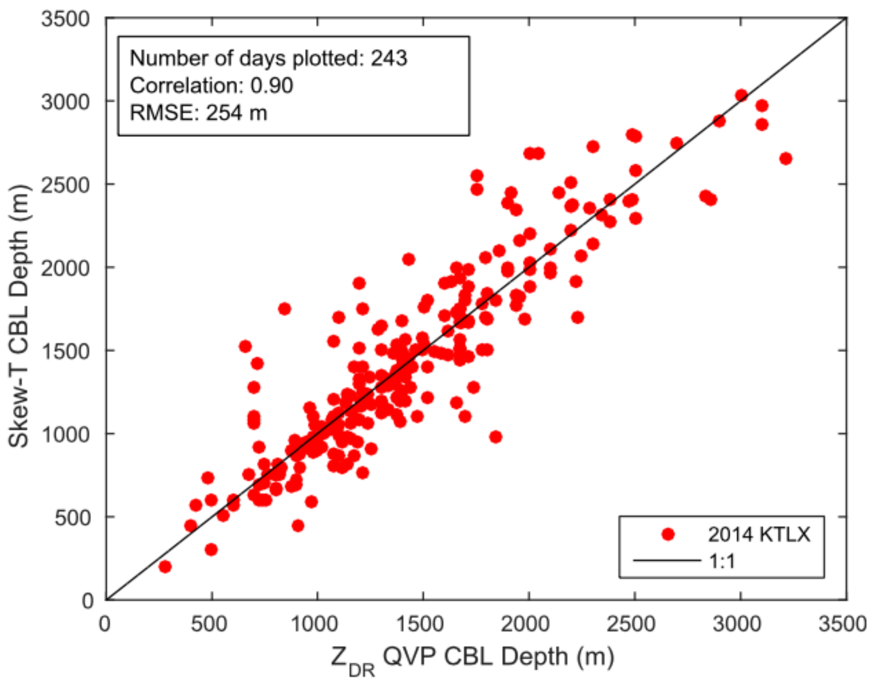


Figure 9 - Comparison of radar and rawinsonde-derived CBL depth estimates for all usable days in 2014. (Reproduction of Fig. 6 from BH18.)

BH18 also tested the robustness of their approach across different temperature and moisture conditions: Minneapolis, Minnesota (KMPX), in February; Fairbanks, Alaska (PAPD), in March; Portland, Oregon (KRTX), in May; Albany, New York (KENX), in June; Tucson, Arizona (KEMX), in August; Riverton, Wyoming (KRIW), in September; Wilmington, Ohio (KILN), in October; and Tampa, Florida (KTBW), in December. The technique applied across the US is shown in Figure 10. Overall, the technique produced a good estimate of CBL depth across all seasons in a variety of environmental conditions. This lends support to the adoption of the approach for all radar sites when compared against other PBLH validation efforts. For example, Bianco et al. (2008) investigated the variability in CBL depth estimation among experts at two locations and found RMSEs of 109 and 135 m. They then used several algorithms to estimate CBL depth with RMSEs between 152 and 424 m depending on the algorithm employed. Similar values of RMSE were found by Elmore et al. (2012).

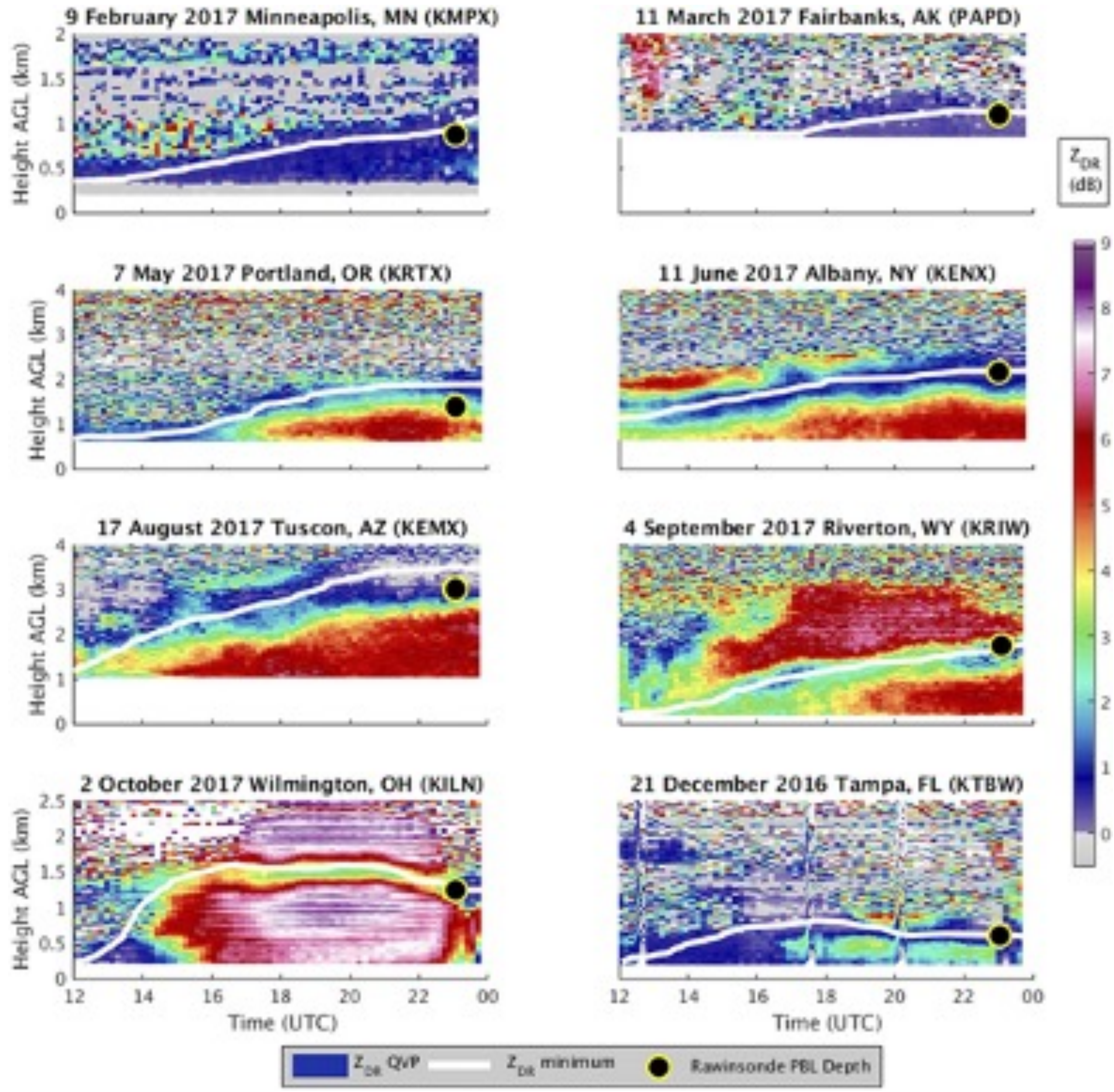


Figure 10: The ZDR QVP on individual days for different locations around the US. The time series of minimum ZDR is manually traced with a white line. Rawinsonde estimates of CBL depth at 2300 UTC are indicated by the black dot outlined in yellow. White shading indicates regions outside the temporal and spatial ranges of radar data. (Reproduction of Fig. 7 from BH18.)

7.2 AER Validation of PBLH Estimates

AER has reviewed the quality of the QVP-based PBLH estimates. The software development phase of the project included a subjective evaluation of the ability of the BH18 quality control and smoothing steps to generate a reasonable PBLH estimate. AER inspected each QVP graph for the approximately 50 days (see Table 2) on which the approach was applied. The diurnal trace of PBLH estimates was compared to scientific expectations. A prominent trough in the ZDR field that rose over time in the mornings was evident on many days, which led to high confidence in the algorithm. On the remaining days, however, patterns of poor height estimates, such as when the automated algorithm selected a ZDR minimum that was unlikely to be correct, were noted. Adjustments were then made to the quality control and smoothing routines, and the cases were rerun to evaluate the corrections.

Table 2 - Dates for which the QVP technique has been applied to radar data

Dates Applied to QVP Technique
20140521-20140522
20140606
20141021
20200709-20200712
20200714-20200715
20210604
20210610
20210613-20210617
20210713
20210718
20210721-20210722
20210726-20210728
20210806
20210812
20210816
20210829-20210903
20210907-20210911
20210921-20210927
20211006-20211008

Objective verification similar in nature to BH18 effort was then performed. TCEQ provided two potential external validation datasets. The sole validation dataset that contains PBLH observation estimates is summarized in Table 3. It is comprised of ozonesonde PBLH estimates coincident in time with NASA flights in the vicitinty of Houston and Galveston, TX, during the summer of 2021. The sites are named Houston, La Porte, Galveston Bay, and Gulf. Unfortunately, the ozonesonde observing campaign was aligned with days for which the radar from 1200 to 0000 UTC was predominantly in precipitation mode. Thus, most of the dates from the ozonesonde dataset are not ideal candidates for use with the PBLH algorithm. No other viable PBLH observations were provided by TCEQ. The NASA TRACER-AQ campaign netCDF files did not contain estimates of PBLH.

Table 3 - Validation data provided by TCEQ

Dates of Interest	Dominant Radar Mode (1200-0000 UTC)	Available NASA Ozonesonde Data			
		Houston	La Porte	Galveston Bay	Gulf
20210604	precip	X			
20210610	missing data			X	
20210613-20210617	mixed	X			
20210713	precip			X	
20210718	precip			X	
20210721-20210722	precip			X	
20210726-20210728	clear then precip	X		X	
20210806	precip	X			
20210812	precip			X	
20210816	precip			X	
20210829-20210903	precip	X	X	X	
20210907-20210911	clear then precip	X	X	X	X
20210921-20210927	mixed then clear	X	X	X	X
20211006-20211008	clear	X		X	

Nevertheless, statistics were generated for all ozonesonde observations provided by TCEQ. The content of the text-format validation file (in meters, where not dimensionless) reflects the overall system performance and is included here:

mbe = 86
rmse = 1089

linear regression:
intercept = 1004
intercept_stderr = 185

pearson correlation coefficient:
r = -0.06
p-value = 0.55

These statistics include days for which the QVP approach is not valid to some extent, but the observations are still included in this analysis because of the overall low count of observations. In practice, the user should be careful to only apply the approach at suitable times on days that are free of major contamination of the ZDR field. A graphical depiction (Figure 11) illustrates the overall poor performance, which is reflective of application on days that are not ideal - here, the majority of days. Of note is the tendency for many PBLH estimates to remain near the minimum height permitted by the algorithm. This behavior should be a focus for improvement once appropriate reliable validation - perhaps based on radiosondes - can be utilized on days with no persistent clouds or precipitation.

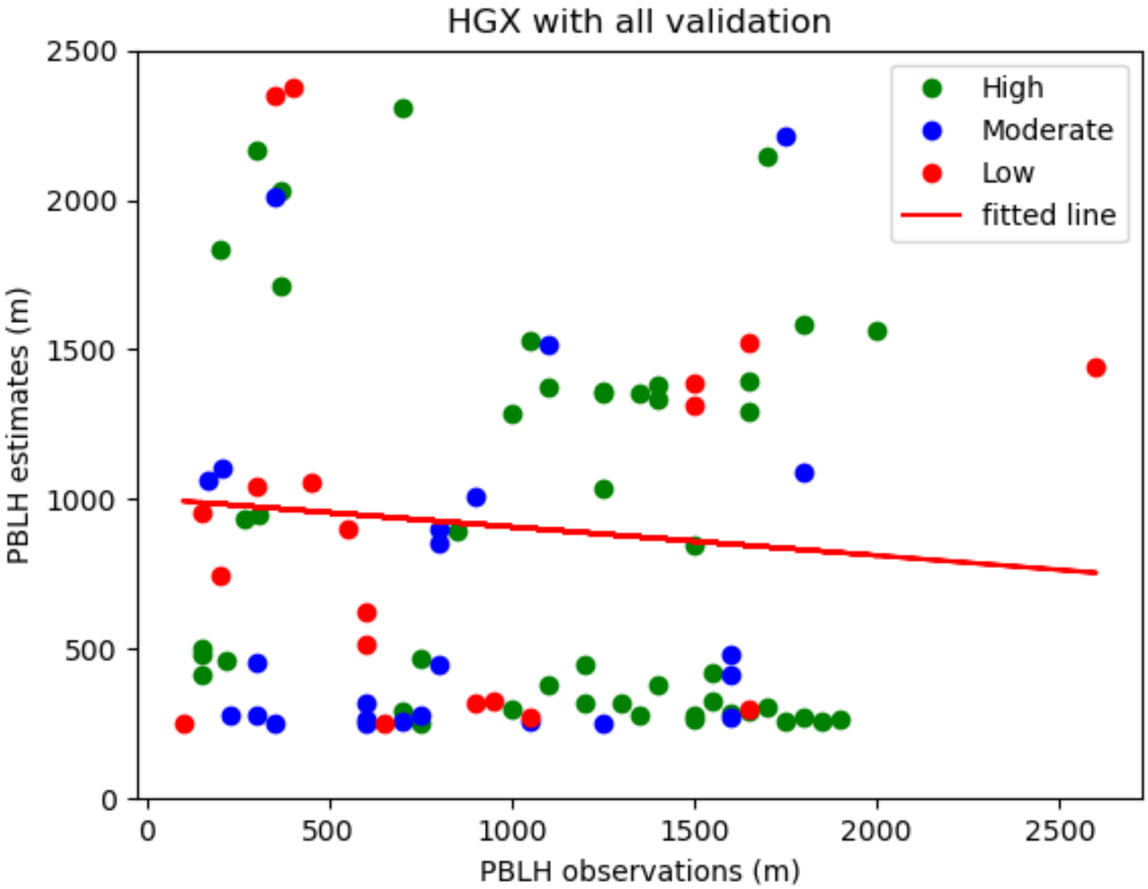


Figure 11 – As in Figure 4, but validation results using all project ozonesonde PBLH observations provided by TCEQ.

Overall, the limited number of quality days for which PBLH validation data were available effectively prevented the generation of useful quantitative validation scores. We anticipate that the technique developed in this project performs similarly to BH18, but we do not have the ability currently to document this fact.

8 Analysis of Project Results

8.1 Major activities and Key Findings

The project developed an end-to-end system for estimating PBL heights from radar data based on the QVP technique used by BH18. Software was developed and installed on the TCEQ computer "eira" as agreed-upon with TCEQ. Training for TCEQ staff enables internal execution of the code and internal adjustments as necessary as the QVP approach is applied to additional days.

Overall, AER believes that the delivered product has great potential as a source of high-quality PBLH estimates close to the radar sites with coverage in and near Texas. The approach can be applied statewide as-is and we recommend that it be run in real-time once a day at the end of the diurnal convective period to demonstrate its utility and to encourage the application of further refinements as needed.

The following key findings highlight our impression of the final product and its usefulness on TCEQ computers:

A - The product will reliably generate PBLH estimates on clear days based on visual inspection of the PBLH time-height plots. Confounding weather conditions, such as clouds, precipitation and any of a variety of other disruptions leading to an inhomogeneous PBL or destruction of the PBL over the region scanned by the radar, will degrade the signal and eventually prevent a PBLH estimate from being retrieved.

B - The technique cannot accommodate every meteorological eventuality and requires ongoing attention and further tuning to ensure proper operation. This can be carried out through inspection of the QVP graphics.

C - Additional validation datasets are needed to provide a quantitative assessment of the technique that is currently in place at TCEQ, however, its performance is anticipated to be similar to the statistics provided by BH18.

8.2 Problems Encountered and Mitigation Strategies

AER scientists note the sensitive nature of the QVP approach when the ZDR field is contaminated. This is not a rare occurrence and is to be expected with this type of radar-based (or any remotely-sensed) approach. Care must therefore be taken to prevent the algorithm from becoming distracted by first understanding the source of the problem. This can be accomplished by close inspection of the QVP graphic on an individual site and case study level. As noted earlier, inhomogeneities in the ZDR field can be caused by innumerable situations. Some are location dependent because of fixed features like mountains, water bodies and wind farms. Each site will suffer from missing or degraded PBLH estimates because of non-optimal meteorological conditions like thick clouds, low clouds or precipitation. Attempts to minimize the occurrence of missing or incorrect PBLH estimates can be accomplished several ways: 1) identify and avoid known poor-quality days (or parts thereof), and 2) refine the QVP algorithm through an iterative procedure involving visual inspection of the QVP plots and subsequent adjustment of the QVP algorithm. Ultimately, the technique is based on sound science and will identify the top of the PBL in the absence of meteorological conditions that can lead the algorithm astray. The user of the code should be aware that visual inspection in the context of the meteorology of the day, subject to awareness of local confounding factors, is always necessary. It should be noted that accommodations for the local factors can, and should be, built into the software.

We contend that the usefulness of the system – in part related to determination of scientific skill and to user confidence - would benefit from validation against a large, standard set of observations for hundreds of cases for each radar site with utility to TCEQ. This will ensure that days (or parts thereof) on which the algorithm does not generate a viable PBLH estimate can be discarded without concern about sample size. Similarly, days for which the radiosonde does not return a reliable PBLH estimate can also be ignored. This approach would increase the speed at which user comfort is established and also encourage adjustments to the scientific routine when patterns of concern are more fully appreciated.

9 Recommendations

Once the above migration strategies have been completed, the system can then address the need to validate NWP model boundary layer heights. Our algorithm inspects the QVP profiles for signatures characteristic of the position of the top of the PBL. The PBL height value that is obtained for each radar scan reflects the position of the top of the boundary layer in a ring typically 20-40 km distant from the radar location, and thus does not represent a single point measurement. In order to use this dataset to evaluate meteorological simulations, an observation operator needs to be developed that will properly average the modeled PBL heights for more direct comparison with the radar values. The next project would develop software to apply the observation operator to WRF output and use it to evaluate a set of WRF simulations performed by TCEQ.

10 References

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.

Elmore, K. L., P. L. Heinselman, and D. J. Stensrud, 2012: Using WSR-88D data and insolation estimates to determine con- vective boundary layer depth. J. Atmos. Oceanic Technol., 29, 581–588, doi: 10.1175/JTECH-D-11-00043.1.

Heistermann, M., S. Collis, M. J. Dixon, S. Giangrande, J. J. Helmus, B. Kelley, J. Koistinen, D. B. Michelson, M. Peura, T. Pfaff, and D. B. Wolff, 2015: The Emergence of Open-Source Software for the Weather Radar Community. Bull. Amer. Meteor. Soc. 96, 117–128, doi: 10.1175/BAMS-D-13-00240.1.

Helmus, J.J. and Collis, S.M., 2016. The Python ARM Radar Toolkit (Py-ART), a Library for Working with Weather Radar Data in the Python Programming Language. Journal of Open Research Software, 4(1), p.e25, doi: 10.5334/jors.119

Ryzhkov, A., P. Zhang, and H. Reeves, 2016: Quasi-Vertical Profiles—A New Way to Look at Polarimetric Radar Data. J. Atmos. Oceanic Technol., 33, 551-562, doi: 10.1175/JTECH-D-15-0020.1