Dow – ARMS -- MSBA

Latest Draft: 3/23/25

Yellow = return to section, see comments/notes

MT DEQ Air Research and Monitoring

2023-2025 Modulair PM2.5 collocation, Analysis and Modeling

Mason Dow

*DRAFT – Format V2*

Table of Contents

[TERMS, ACRONYMS AND ABBREVIATIONS 4](#_Toc193716481)

[EXECUTIVE SUMMARY 5](#_Toc193716482)

[BACKGROUND 6](#_Toc193716483)

[USE 6](#_Toc193716484)

[INTRODUCTION 6](#_Toc193716485)

[MT DEQ AIR RESEARCH AND MONITORING 6](#_Toc193716486)

[OVERVIEW 6](#_Toc193716487)

[POLLUTANTS 6](#_Toc193716488)

[AQI 7](#_Toc193716489)

[SAMPLERS/MONITORS/SENSORS 7](#_Toc193716490)

[UMT MSBA CAPTSONE 8](#_Toc193716491)

[OVERVIEW 8](#_Toc193716492)

[GITHUB REPOSITORY 8](#_Toc193716493)

[CODE 8](#_Toc193716494)

[PROBLEM STATEMENT 9](#_Toc193716495)

[DATA COLLECTION 9](#_Toc193716496)

[OVERVIEW 9](#_Toc193716497)

[COLOCATION 9](#_Toc193716498)

[SITE SELECTION 9](#_Toc193716499)

[PHYSICAL COLOCATION 10](#_Toc193716500)

[QUALITY CONTROL AND ASSURANCE 11](#_Toc193716501)

[DATA QUERYING 11](#_Toc193716502)

[VARIABLES INCLUDED 12](#_Toc193716503)

[EXPLORATORY DATA ANALYSIS 13](#_Toc193716504)

[PREPROCESSING 13](#_Toc193716505)

[ANALYSIS TRANSFORMATIONS 13](#_Toc193716506)

[DESCRPTIVE STATISTICS 13](#_Toc193716507)

[DATA VISUALIZATIONS 15](#_Toc193716508)

[BOXPLOTS 15](#_Toc193716509)

[TIME SERIES LINE GRAPHS 17](#_Toc193716510)

[LINES OF BEST FIT 18](#_Toc193716511)

[MEASURES OF MODULAIR ACCURACY 20](#_Toc193716512)

[MODELING 20](#_Toc193716513)

[PURPLEAIR EPA CF1 20](#_Toc193716514)

[DIFFERENCES/SIMILARITIES 21](#_Toc193716515)

[MODELING PROCESS 21](#_Toc193716516)

[EXCLUSION OF NC, LT SITES 21](#_Toc193716517)

[TIDYMODELS PROCESS 21](#_Toc193716518)

[TESTING/TRAINING SPLITS 21](#_Toc193716519)

[MODELS TESTED 21](#_Toc193716520)

[RESULTS 21](#_Toc193716521)

[DATA VISUALZATIONS BY MODEL 21](#_Toc193716522)

[UPDATED MEASURES OF ACCURACY BY MODEL 21](#_Toc193716523)

[MODEL SELECTION 21](#_Toc193716524)

[IMPLEMENTATION 21](#_Toc193716525)

[LIMITATIONS 21](#_Toc193716526)

[CONCLUSIONS 21](#_Toc193716527)

[NEXT STEPS 22](#_Toc193716528)

# TERMS, ACRONYMS AND ABBREVIATIONS

|  |  |
| --- | --- |
| MT | Montana |
| DEQ | Department of Environmental Quality |
| AQB | Air Quality Bureau |
| ARMS | Air Research and Monitoring Section |
| EPA | Environmental Protection Agency |
| PM2.5 | Particulate matter of 2.5 microns or less in diameter |
| PM10 | Particulate matter of 2.5 microns or less in diameter |
| UG/M3 | Micrograms per cubic meter |
| CO | Carbon monoxide |
| NO | Nitrogen monoxide |
| NO2 | Nitrogen dioxide |
| O3 | Ozone |
| BAM | Beta attenuation monitor |
| FEM | Federal Equivalency Method |
| CFR | Code of Federal Regulations |
| SLAMS | State or Local Air Monitoring Stations (regulatory grade) |
| AQI | Air Quality Index |
| CAA | Clean Air Act |
| NAAQS | National Ambient Air Quality Standards |
| Ambient air monitors | May be referred to as “monitors”, “samplers”, “sensors,” or “BAMs” (context dependent) |
| UMT | The University of Montana |
| MSBA | Master of Science of Business Analytics |

# EXECUTIVE SUMMARY

To be completed at end

# BACKGROUND

## USE

This report is primarily intended for internal use by MT DEQ Air Quality Bureau staff. It is intended to serve as both a high-level summary of analysis and findings, as well as a technical document enabling reproducibility of work and assuring quality control of processes described within. Additional background information is included for audiences unfamiliar with MT DEQ ARMS goals and procedure.

All additional questions, comments and concerns can be forwarded to the project’s author, Mason Dow: [mason.dow@mt.gov](mailto:mason.dow@mt.gov)

## INTRODUCTION

This report summarizes the purpose, methods, analysis, modeling and correction of Montana Department of Environmental Quality Air Research and Monitoring Section operated QuantAQ Modulair PM2.5 air quality sensors, collocated with regulatory-grade FEM PM2.5 ambient air monitors. All findings herein are preliminary and meant for internal use only. This report is meant to cover an initial period of data collocation and analysis between the regulatory grade PM2.5 samplers and the Modulair units for a single pollutant. This project is ongoing and will continue, analyzing a greater variety of pollutants, samplers, and concentration levels in subsequent efforts.

## MT DEQ AIR RESEARCH AND MONITORING

### OVERVIEW

The Air Research and Monitoring Section (ARMS) operates an ambient air monitoring network across the state of Montana. The purpose of Montana’s ambient air monitoring network is to monitor, assess and provide information on statewide air quality conditions and trends as specified by the Montana and Federal Clean Air Acts. The Air Quality Monitoring Program works in conjunction with local air pollution agencies and some industries, measuring air quality throughout Montana. This data provides the factual basis for regulatory decisions as well as provides air quality information to our local counterparts and the public.

In additional to its regulatory ambient air monitors, ARMS operates several informational air monitors for a variety of purposes related to its overall mission. This includes public information and ancillary scientific studies. Some of these samplers are regulatory-quality instruments modified to exist as special purpose monitors (SPMs). Others are light-scattering sensors. (need more + transition here)

### POLLUTANTS

PM2.5, or particulate matter of 2.5 microns or less in diameters, is Montana’s “pollutant of concern”, or a key focus of the ambient air monitoring network. Particulate matter is one of the six [Criteria Air Pollutants](https://www.epa.gov/criteria-air-pollutants/information-pollutant) as defined by the National Ambient Air Quality Standards (NAAQS), determined by the EPA. Wood smoke, from wildfires and wood stoves, is the chief source of PM2.5 across the state. Monitoring this pollutant is of great important to ARMS. Increased focus on recent impactful wildfire smoke seasons (primarily the summer months, from May to September), coupled with recently lowered national ambient air quality standards for PM2.5, citizens and local governments and seeking expanded PM2.5 monitoring and information for making sound decisions for the benefit of public health. Both regulatory and non-regulatory monitors are employed to provide the public with PM2.5 concentrations for airsheds across the state.

### AQI

ARMS seeks to provide the public with air quality information in real-time. Hourly measurements of PM2.5 are recorded by the monitors and reported out. Data is available on the DEQ-hosted [Today’s Air dashboard](https://todaysair.mtdeq.us/), as well as the [EPA’s AirNow](https://www.airnow.gov/) air quality portal and various tools. This data is communicated to the public using two key units; UG/M3 (micrograms per cubic meter), which is a concentration (mass per volume), and the Air Quality Index (AQI). The U.S AQI was developed by the EPA for communicating information about outdoor air quality and health. The AQI includes six color-coded categories, each corresponding to a range of index values. The higher the AQI value, the greater the level of air pollution and the greater the health concern. For example, an AQI value of 50 or below represents good air quality, while an AQI value over 300 represents hazardous air quality, while 100 generally corresponds to an ambient air concentration that equals the level of the short-term national ambient air quality standard (NAAQS) for protection of public health. Each category corresponds to a different level of health concern and has a specific color, making it easy for people to quickly determine whether air quality is reaching unhealthy levels in their communities. There exist AQI values for five of the major air pollutants regulated by the Clean Air Act (CAA). For purposes of this report, all AQI values refer to those corresponding to PM2.5 concentrations (fig. A).

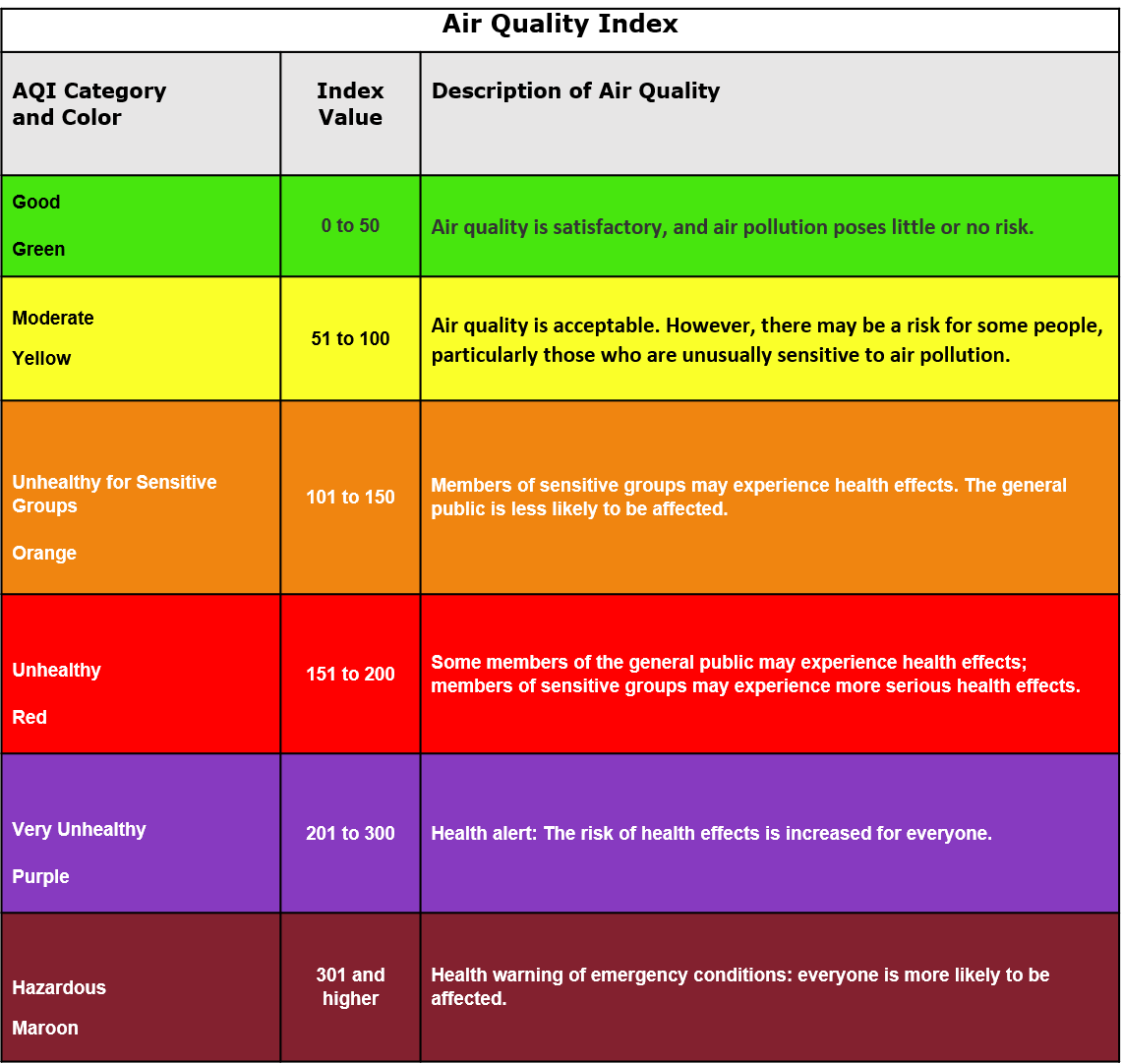


Figure : EPA Air Quality Index Table (source: AirNow.gov)

An additional note: The EPA also calculates an [AQI NowCast value](https://www.epa.gov/sites/default/files/2018-01/documents/nowcastfactsheet.pdf), which uses a weighted averaging system of concentrations to estimate a more “current” air quality value. This report and its analysis does not use the NowCast averaging system, instead opting to use the traditional static AQI values for simplicity.

### SAMPLERS/MONITORS/SENSORS

Two broad categories of air samplers have been used for this project.

One is the regulatory grade Federal Equivalency Method (FEM) Beta Attenuation Monitors (BAM). Federal Equivalency Method refers to the [Code of Federal Regulations](https://www.ecfr.gov/current/title-40/chapter-I/subchapter-C/part-50/appendix-Appendix%20L%20to%20Part%2050'), which governs the processes and equipment by which air programs like DEQ ARMS can produce air quality data that can be used in regulatory decisions, such as an airshed being in our out of attainment with a given NAAQS level (for PM2.5 there are annual and daily averages which when exceeded incur regulatory action by DEQ and the EPA). FEM samplers are determined to be equivalent, over a twenty-four-hour sampling period, to the Federal Reference Method. Beta attenuation uses a radioactive carbon source and detector, passed through teflon or nylon filter tape, with detector measurements made before and after air sampling. Sampling draws a steady, defined flow of ambient air over the filter tape. How much radioactive decay that is blocked from reaching the detector corresponds to a mass of particulate matter present on the filter tape. As such, a concentration of particulate matter in the ambient air over that hour can be determined.

The FEM BAMs used for this effort included the 1020 BAM and 1022 BAM, manufactured by MetOne Instruments, and the 5014i (also known to ARMS as a Thermo), manufactured by Thermo Fisher Scientific.

The second are air sensors. Those employed are non-regulatory air quality measuring instruments. These use the principles of light scattering and a light detector to estimate the concentration of suspended particulate matter in the air. They can estimate concentrations at much more frequent intervals due to the nature of their measurement techniques.

The air sensors used for this project are twofold. The first is the primary instrument of concern, the QuantAQ Modulair (Modulair) sensor. The second is the PurpleAir Classic plus, which is collocated in all the same sites where Modulair units have been collocated and is used for comparison purposes.

## UMT MSBA CAPTSONE

### OVERVIEW

This project has been integrated as part of my Capstone, completing my Master of Science of Business Analytics. All products and processes have been operated and built with the client’s needs, the ARMS section, as the priority of the effort.

### GITHUB REPOSITORY

The code for completing the data retrieval, analysis, modeling and report outputs for this project are stored in a public cloud GitHub repository, owned by me, found here: <https://github.com/masondow/mt_deq_arms_modulair_colocation>

The README file for the repository contains all information related to the process for reproducing the analysis and modeling used for this analysis. The README contains an Overview, Objectives, Repository Structure and Executables sections. Raw data is ultimately withheld from the repository, following best practices.

### CODE

Code is well commented and conforms to [Google’s style guide for Python](https://google.github.io/styleguide/pyguide.html) (which is slightly more elaborate PEP 8 style), and the [Tidyverse style guide for R](https://style.tidyverse.org/). Function files are used and called in for built functions in both R and Python. R code for exploratory data analysis and modeling is achieved via R Markdown format, more on which can be found [here](https://rmarkdown.rstudio.com/).

Code is written to be as modular and adaptable as possible for future use, primarily with considerations for inclusions of other pollutants measured by Modulair units and use by other ARMS staff.

# PROBLEM STATEMENT

ARMS continually seeks to expand its ambient monitoring network for the benefit of public health and air quality information. This effort is constrained by the staffing time and financial resources. The cost of regulatory-grade ambient air monitoring equipment is high in both, given the substantial price tag for samplers and the extensive quality control, quality assurance, maintenance and installation labor required for the operation.

As such, ARMS looks to non-regulatory grade samplers, sensors, which can lower the time and financial investment required to produce air quality data. However, research has demonstrated that various models of air sensors can have significant bias and inaccuracy when compared to regulatory grade equipment. ARMS seeks to hold its air quality data production for public consumption to the highest standard, given its importance in empowering Montana’s to make sound decisions regarding their health and wellbeing. Thus, ensuring the accuracy of all non-regulatory grade samplers is paramount.

This project seeks to analyze, and correct, the accuracy of a new air sensor, the QuantAQ Modulair sensor. QuantAQ is a newer company founded in 2021, and their Modulair unit presents an exciting opportunity for air programs. The Modulair measures both a range of particulate matter size concentrations (PM1, 2.5, 10), as well as four gas-phase pollutants; carbon monoxide (CO), oxides of nitrogen (NO and NO2), and ozone (O3). This sampling capability in a single instrument, if found reliable, would be a significant boon to expanding ambient air monitoring efforts across Montana’s vast geography.

# DATA COLLECTION

## OVERVIEW

Assessing and correcting for the accuracy of the Modulair takes a substantial dataset. Chiefly, Modulair units need to be collocated with FEM instruments. The physical collocation process needs to be sound and follow best practices. The data produced by both instruments at a collocated site needs to be quality controlled and assured to all necessary standards. Finally, the data needs to be queried and preprocessed for analysis and modeling.

## COLOCATION

### SITE SELECTION

Modulair colocation and sampler installation at existing ARMS FEM PM2.5 sites was done in a stepwise fashion due to purchasing and inventorying of the Modulair units themselves, configuration of installation hardware, and integration with ARMS database software.

The following sites in Table 1 were chosen for Modulair colocations:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Site** | **Site ID** | **Installation Date** | **Site Type** | **City** |
| Rossiter Elementary School | RP | 11/16/2023 | Current SLAMS site | Helena |
| NCORE | NC | 11/21/2024 | Current SLAMS site  (background) | Helena |
| Lewistown | LT | 12/12/2024 | Current SLAMS site | Lewistown |
| Missoula Boyd Park | MP | 2/18/2025 | Current SLAMS site | Missoula |
| Libby Courthouse Annex | LB | 2/19/2025 | Current SLAMS site | Libby |
| Billings Lockwood | BW | 3/5/2025 | Current SLAMS site | Billings |

Table : Modulair Colocation Sites

These sites were chosen for their proximity to ARMS offices in Helena, and historical likelihood for receiving elevated wintertime PM2.5 levels, given the need to observe as wide a range as possible of concentrations to avoid less complete analysis or overfit models. Green shaded rows represent sites that were determined to be of use for inclusion in this projects analysis and modeling; the other three sites will be included for PM2.5 analysis and modeling in future efforts.

The Rossiter (site ID: RP) site is the longest-running Modulair collocated unit, installed in November of 2023. As such, the vast majority of collocation data comes from this site. Unique to the Rossiter site is the fact that it has three BAMs located at the site, allowing for a cross-sampler comparison to the Modulair unit installed there. See Analysis for more.

### PHYSICAL COLOCATION

All colocation of both the Modulair sensor and the PurpleAir sensors follow best practices for siting and installation of collocated sensors as [recommended by the EPA](https://www.epa.gov/air-sensor-toolbox/air-sensor-collocation). All sites are SLAMS sites and meet all EPA siting criteria.



Figure : The Modulair colocation at the NCORE site, an example of excellent sensor siting

## QUALITY CONTROL AND ASSURANCE

FEM samplers at all sites for this project are operated to regulatory grade CFR standards for monitor operation. These standards include running the samplers within a specified flow right, in appropriate ambient conditions, and using all pertinent EPA-approved ancillary sampling equipment like cyclones and downtubes.

Data is quality controlled to CFR standards and is nulled when these standards are not met. This could include the ambient temperature within a site shelter falling out of range for the BAM manufacturer’s recommendation, or a leak or flow audit failing during a monthly quality control check, wherein data would be nullified to the last passing check.

Data from the Modulair units is monitored in real-time, but not audited to the same standards as an FEM BAM given the differences in equipment and use. Flow rates from the small sensor fan are not audited, as done with BAM pumps. Suspect data was removed from the dataset. See Preprocessing for more.

## DATA QUERYING

Several distinct processes were used to query data for this project on my local machines. ARMS data is stored in a Microsoft SQL Server-backed relational database, configured using a bespoke environmental data management software called AirVision, developed by Agilaire. A custom R package called for querying this database (called AVconn) has been developed by ARMS staff and has been employed for prior iterations of analysis for collocated sensor data. That package was ignored for this effort given a need for a few key updates to the package for querying non-regulatory formatted data like the Modulair.

FEM data was downloaded as Excel files onto my local machine using the AirVision software GUI. This was the easiest path for pulling this type of data, and exported the data in ready-to-use formats for merging with the other datasets (i.e. the datetime column formatted to YYYY-MM-DD HH:MM:SS).

Modulair data was queried from the [QuantAQ Cloud API](https://docs.quant-aq.com/software-apis-and-libraries/quantaq-cloud-api) using a custom Python script and a private API key. One’s own API key must be obtained to query this database. Data is passed via JSON files. The QuantAQ Cloud API is well documented and simple to navigate. While ARMS-owned Modulair data does currently exist in the AirVision database, as scheduled API polling and file importation between the two cloud databases was configured in 2024, the AirVision-stored dataset for the Rossiter unit is incomplete due to early difficulties encountered when attempting to back-poll the QuantAQ database for older data. As such, querying via the API was the easier solution. While minute-average data can be retrieved from the Modulair units, hourly averaged data (resampled data) was queried to skip re-averaging the data to match the BAMs’ sampling interval. This data was converted, date-formatted, column-renamed and saved separate excel files by site.

PurpleAir data was queried from a separate ARMS-owned SQLite database and underwent the same data transformations in Excel.

Data from the three sources was then joined on the date in Excel using PowerQuery. Files were separated by site. Table 2 shows the example one-dimensional file and the column naming structure used for the NCORE site, as well as the variables included in querying:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **siteid** | **datetime** | **siteid\_datetime** | **nc\_1020** | **nc\_modulair\_pm25** | **nc\_modulair\_rh** | **nc\_modulair\_tempc** | **nc\_purpleair\_cf1** | |
| NC | 2024-11-22 08:00:00 | NC\_2024-11-22 08:00:00 | 6.5 | 4.048 | 67.903 | -3.403 | 10.139497 |
| NC | 2024-11-22 09:00:00 | NC\_2024-11-22 09:00:00 | 6.8 | 8.603 | 68.258 | -3.463 | 11.759419 |
| NC | 2024-11-22 10:00:00 | NC\_2024-11-22 10:00:00 | 1.1 | 5.778 | 69.228 | -3.717 | 11.0322113 |
| NC | 2024-11-22 11:00:00 | NC\_2024-11-22 11:00:00 | 1.6 | 5.171 | 70.05 | -3.802 | 10.67521 |
| NC | 2024-11-22 12:00:00 | NC\_2024-11-22 12:00:00 | 1.4 | 6.048 | 68.922 | -3.857 | 10.9713253 |
| NC | 2024-11-22 13:00:00 | NC\_2024-11-22 13:00:00 | 1.6 | 7.042 | 67.018 | -3.565 | 10.6620331 |

Table : Sample table head from final file format for the NCORE site

### VARIABLES INCLUDED

* Datetime (YYYY-MM-DD HH:MM:SS)
* SiteID
* Concatenated datetime & siteID
* All PM2.5 measurements
  + Formatted: [siteID]\_[sampler]\_[pm25]
  + 3 FEM measurements at RP
  + PurpleAir PM2.5 measurement is using the EPA correction factor, cf1
* Ambient temp (degC) recorded by Modulair
* Relative humidity (RH %) recorded by Modulair

# EXPLORATORY DATA ANALYSIS

## PREPROCESSING

* FEM data will Null Codes removed
* PurpleAir data with null codes removed (null code derived from secondary QA procedure ensuring channel alignment, not present with Modulair units)
* Suspect sensor data (>500 ug/m3, >|100| ug/m3 difference from FEM measurement) nulled in dataset

## ANALYSIS TRANSFORMATIONS

* FEM average at RP site
  + Exploration of weighted average but abandoned
* Built corresponding numeric AQI categories for all PM2.5 measurements
* Transformed to long Tibbles for graphing

## DESCRPTIVE STATISTICS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Descriptive Statistics Table – by Site | | | | |
| Parameter | **Mean** | **Standard Deviation** | **Min** | **Max** |
| rp\_1020 | 9.488266 | 10.673886 | -10.000000 | 114.70000 |
| rp\_1022 | 7.010990 | 7.438765 | -10.000000 | 62.80000 |
| rp\_thermo | 7.927400 | 8.524063 | -9.600000 | 111.40000 |
| rp\_fem\_avg | 8.254254 | 8.521646 | -10.000000 | 104.55000 |
| rp\_fem\_weighted\_avg | 6.581120 | 6.931242 | -5.557163 | 70.39881 |
| rp\_modulair\_pm25 | 8.099645 | 11.036336 | 0.051000 | 169.82300 |
| rp\_modulair\_rh | 54.088105 | 20.972775 | 7.453000 | 94.65000 |
| rp\_modulair\_tempc | 6.909123 | 13.026512 | -39.837000 | 43.43800 |
| rp\_purpleair\_cf1 | 7.184004 | 8.783683 | -2.512187 | 154.32125 |
| nc\_1020 | 1.884376 | 4.790030 | -10.0000000 | 57.20000 |
| nc\_modulair\_pm25 | 5.235424 | 9.579881 | 0.0240000 | 78.60200 |
| nc\_modulair\_rh | 59.829025 | 14.352805 | 22.0030000 | 91.36500 |
| nc\_modulair\_tempc | -2.176947 | 8.122159 | -29.4820000 | 16.30500 |
| nc\_purpleair\_cf1 | 4.156230 | 3.541788 | -0.5153173 | 26.96663 |
| lt\_thermo | 2.8792208 | 3.872616 | -10.100000 | 26.0000 |
| lt\_modulair\_pm25 | 0.9101154 | 1.449905 | 0.001000 | 23.1320 |
| lt\_modulair\_rh | 60.6199078 | 16.762165 | 16.150000 | 89.7570 |
| lt\_modulair\_tempc | -3.3591532 | 9.380643 | -29.758000 | 19.4170 |
| lt\_purpleair\_cf1 | 2.4741148 | 2.993761 | -2.042637 | 43.4133 |

Table : Descriptive statistics by site and parameter

Takeaways:

## DATA VISUALIZATIONS

### BOXPLOTS

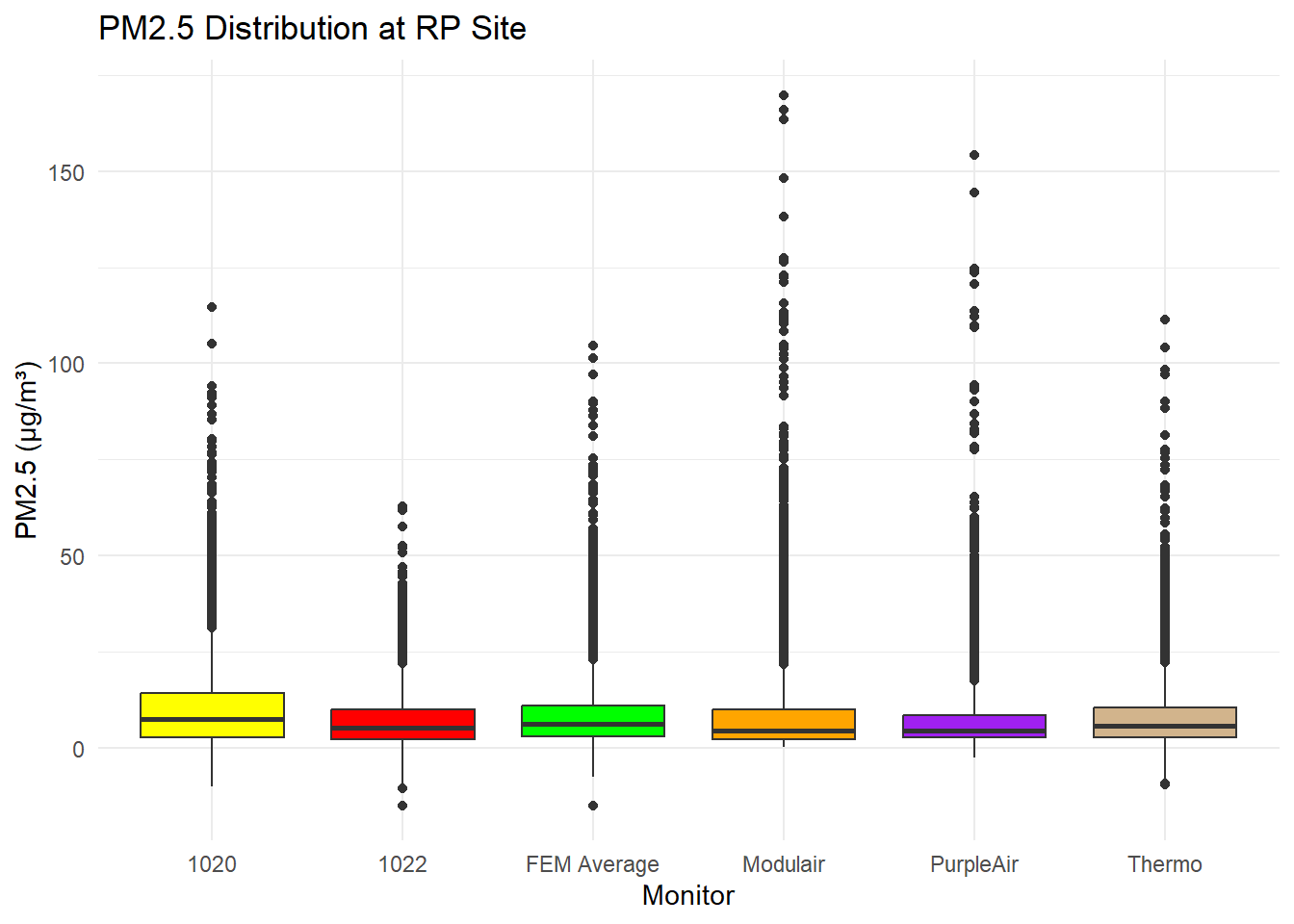


Figure :

Chart, box and whisker chart

Description automatically generated

Figure :

Chart, box and whisker chart

Description automatically generated

Figure :

### TIME SERIES LINE GRAPHS

Messy, will probably shorten to single week window or calculate **daily** averages

Chart, histogram

Description automatically generated

Figure :

Chart, histogram

Description automatically generated

Figure :

Chart, histogram

Description automatically generated

Figure :

### LINES OF BEST FIT

Chart, scatter chart

Description automatically generated

Figure :

Chart, scatter chart

Description automatically generated

Figure :

Chart, scatter chart

Description automatically generated

Figure :

# MEASURES OF MODULAIR ACCURACY

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Overall Accuracy Metrics by Site (FEM vs. Modulair) | | | | | | | | |
| **Site** | **R\_squared** | **Average\_Error** | **MAE** | **RMSE** | **NRMSE** | **Slope** | **Intercept** | **aqi\_match\_frequency** |
| rp | 0.6693856 | 0.1671611 | 3.833668 | 6.360686 | 76.95288 | 1.0571365 | -0.6394333 | 0.8240882 |
| nc | 0.0172710 | -2.3103267 | 4.926509 | 8.997140 | 475.76500 | 0.2160249 | 3.7928934 | 0.8540907 |
| lt | 0.0285074 | 1.9674870 | 3.204496 | 4.366840 | 151.66742 | 0.0633016 | 0.7294745 | 0.9383117 |

Table :

Need to re-do Accuracy Metrics by AQI Category by Site Tables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Accuracy Metrics by AQI Category – All Sites | | | | | | | |
| **AQI category** | **Count** | **R\_squared** | **MAE** | **Average\_Error** | **RMSE** | **NRMSE** |
| 1 | 11411 | 0.0558350 | 2.931397 | -0.2248953 | 5.070934 | 158.79801 |
| 2 | 3694 | 0.4144304 | 6.022049 | 1.4531934 | 7.935956 | 51.37488 |
| 3 | 148 | 0.1942612 | 16.259795 | -7.3921914 | 20.876425 | 49.99883 |
| 4 | 34 | 0.1778051 | 35.431500 | -19.7313824 | 42.661724 | 59.89217 |

Table :

Takeaways:

# MODELING

## PURPLEAIR EPA CF1

Explain CF1 and EPA Modeling/Correction effort

### DIFFERENCES/SIMILARITIES

## MODELING PROCESS

### EXCLUSION OF NC, LT SITES

### TIDYMODELS PROCESS

### TESTING/TRAINING SPLITS

### MODELS TESTED

## RESULTS

### DATA VISUALZATIONS BY MODEL

### UPDATED MEASURES OF ACCURACY BY MODEL

### MODEL SELECTION

## IMPLEMENTATION

## LIMITATIONS

# CONCLUSIONS

# NEXT STEPS