CFARS SS Subgroup Phase 2 Analysis Calculation Tool Tutorial

For Mac and Windows

August 9, 2019

Glossary

- Opensource – computer program in which the source/ master code is available to the public for use and/or public and or private modification.

- Github - software development platform that allows developers to discover, share, and build software (e.g., code). It is a web-based hosting service for version control using Git.

- Github Repository – Similar to a dedicated project folder. Your project's repository contains all of your project's files. For collaboration, repositories are often shared.

- Anaconda – Free and opensource distribution of Python and R programming languages for data science. Anaconda is primarily used for program package management and deployment.

- Python – a programming language.

Summary of Major Steps for Phase 2:

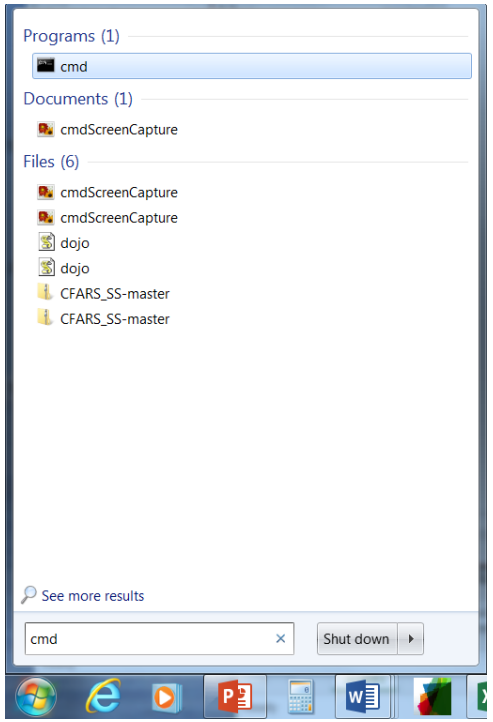
- **Step 1:** Get/update the code
- **Step 2:** prepare your data set
- **Step 3:** update the configuration template
- **Step 4:** Run the code for a single dataset
- **Step 5:** Check your project results matrix
- **Step 6:** Repeat Steps 2-5 for each project dataset
- **Step 7:** Submit your Phase 2 analysis result matrices

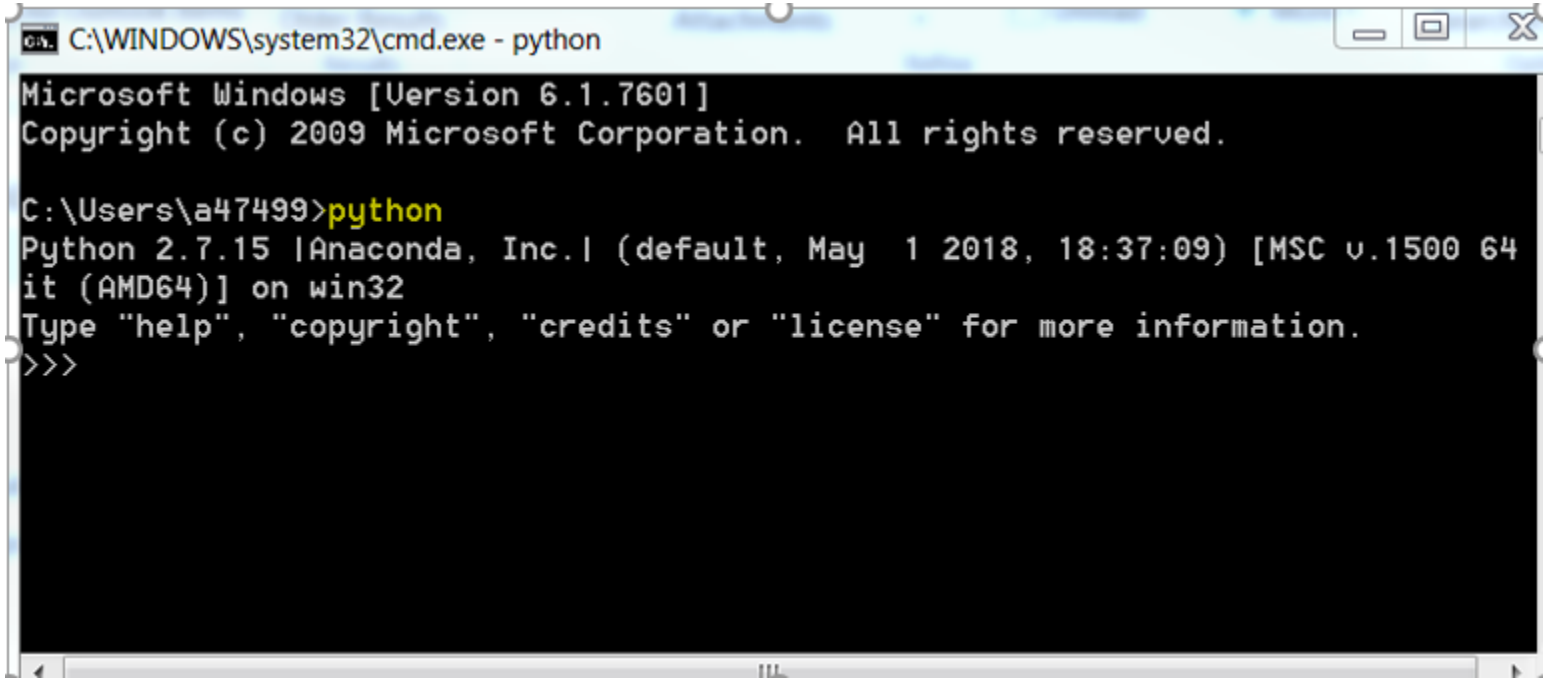
Support Documentation

- CFARS_SS_Subgroup_Phase_1_Analysis_Tool_Tutorial_v04.pdf
- YouTube tutorial for phase1: https://youtu.be/wtyHk7J1NV4
- Example_Phase1Test_Output_20181203
- *CFARS_SS_Subgroup_Phase_2_Analysis_Tool_Tutorial.pdf
- *Example_Phase2Test_Output_20190809

Quick check: python Installed?

- The tutorial assumes the necessary software was installed during phase 1. Confirm python is installed by navigating to your start window, type cmd, to access windows command window. Type “python” in the command window.



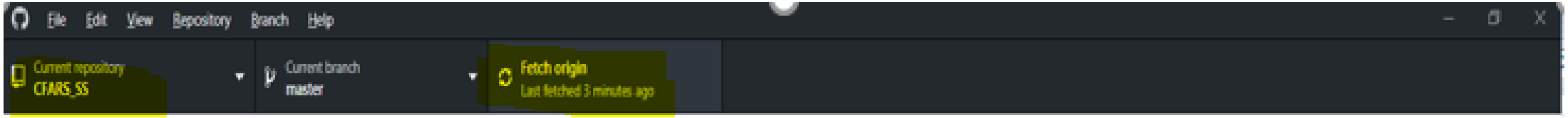


- If you receive an error message, please refer to the phase 1 tutorial that covers installing the necessary software. NOTE: PYTHON 2 or 3 will work

Step 1: Get/update the code

If you ran the phase 1 tool on your computer, you should already have a directory on your computer called **CFARS_SS**. For phase 2, all you will need to do is sync that directory to get the updated code

- if you are familar with git, go ahead and pull the new and updated repository with the "git pull" command in you terminal or command prompt from the CFARS_SS folder on your system
- OR in git hub desktop navigate to the CFARS_SS repository and click the "Fetch origin button" in the top ribbon of the repository view

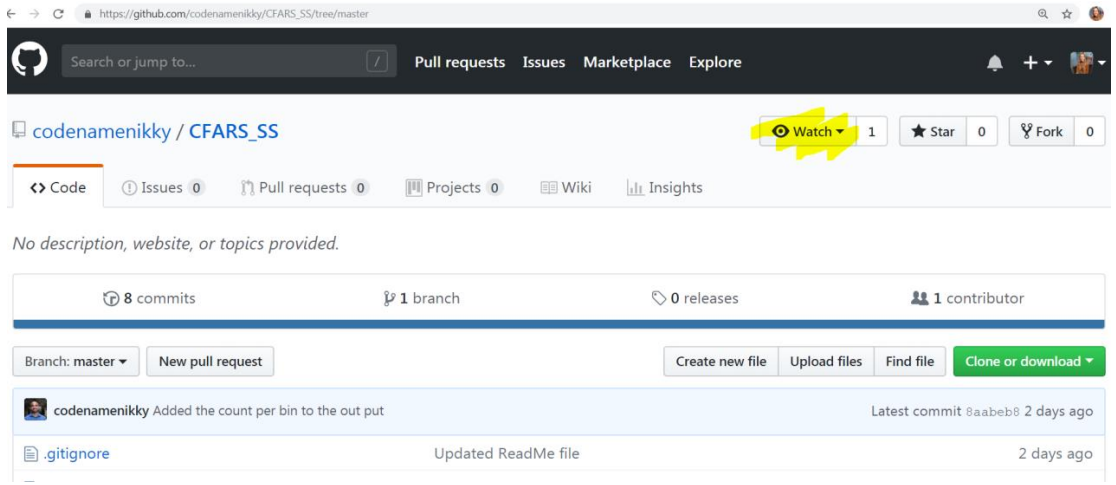


VERY IMPORTANT: Select “fetch origin” prior to running the code as you want to make sure the master code on your computer is latest version of open source master code. Think of it as a code “refresher” button. If there are changes, it will ask you if you want to "pull origin", and you should accept the option to do that. You will know that your repository is updated if you see CFARS_SS_Phase2_Analysis.py in your CFARS directory.

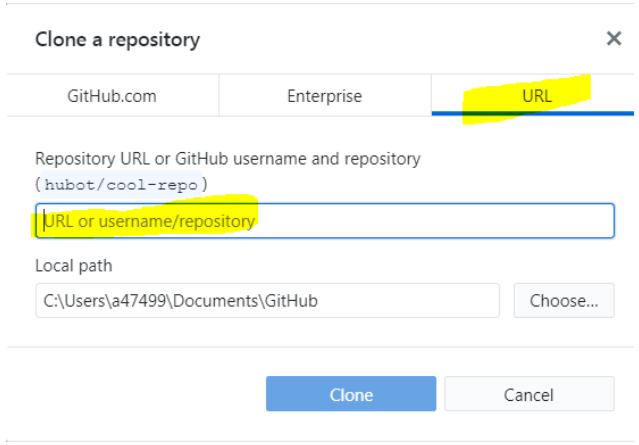
Step 1: Get/update the code

Scenario: You don't have the repository on your computer (the CFARS_SS directory) and or do not have git or git desktop

- Refer to phase 1 tutorial to download Git or Git desktop
- Navigate to the CFARS SS Subgroup Github repository page online
  - https://github.com/codenamenikky/CFARS_SS/tree/master
  - The page should automatically log you in… (place cursor over top right corner avatar to confirm this)
  - This is the opensource page that shows the master code and where all the updates and code issues are reported

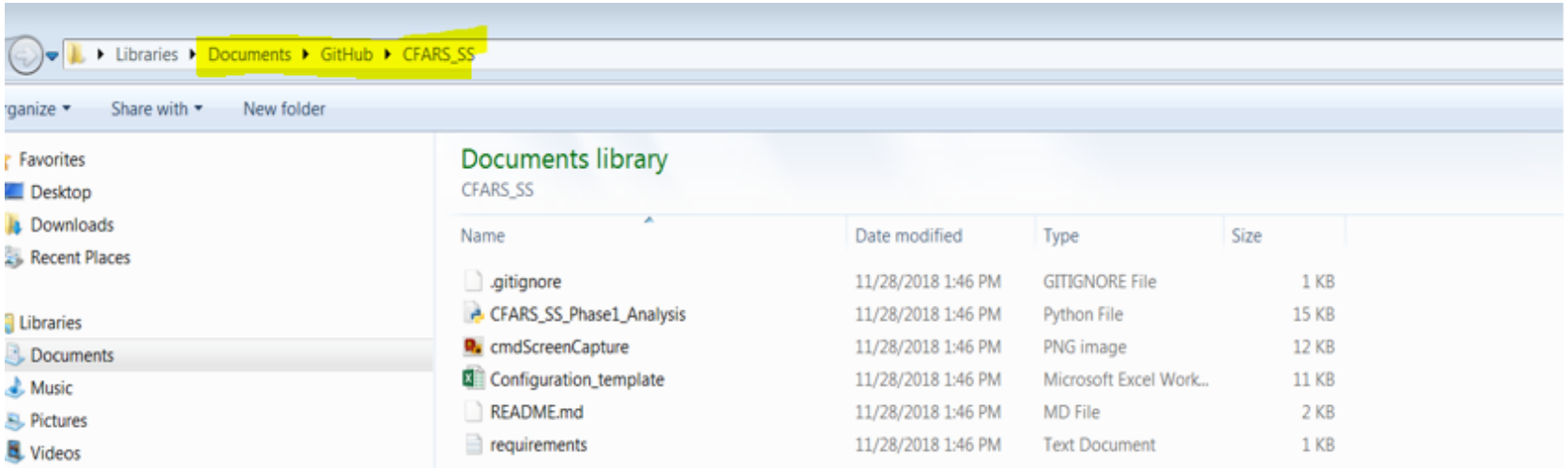


- On the Github repository, select on “Clone or download” to save the master code and required input files to your computer’s Github Desktop interface
- Open Github Desktop
- Copy the weblink and paste in your Github Desktop application directly
  - File →Clone Repository →URL



Step 1: Get/update the code

- Next, confirm the required folders have been added to your computer's "Documents" folder
  - Github → CFARS_SS
    - All of the items inside this folder are synced with the group's opensource master code.

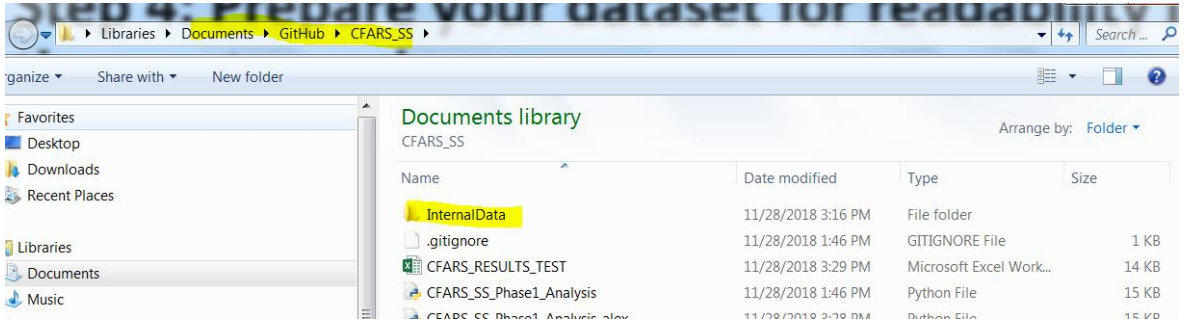


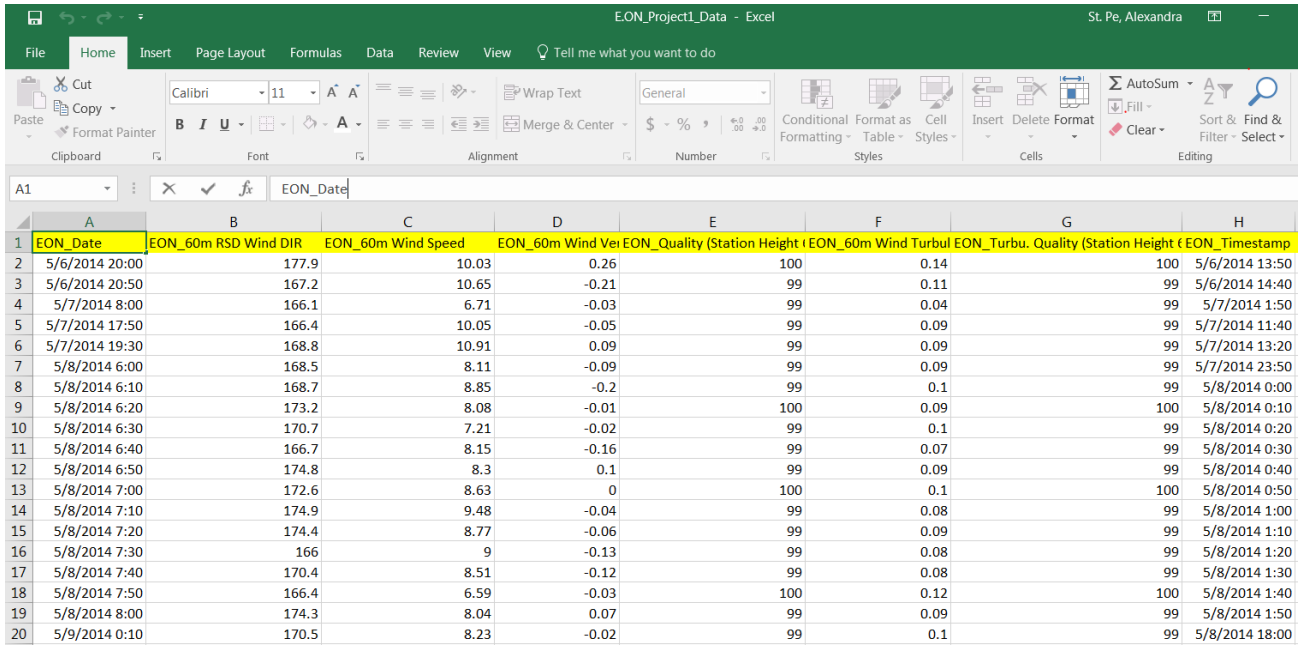
Step 2: Prepare your dataset for readability in the opensource master code

- The master code provides TI bias stats one project at a time. So each project datasets must be a separate file

- Create a folder named “InternalData” and save it in your CFARS_SS folder

- Next, modify your project datasets so they each have only **one row of header information (see example below)**
  - You *must* save each one header project dataset file in Excel OR as a .csv format for the master code to read it
- Finally, save your modified project datasets in the “InternalData” folder





	A	B	C	D	E	F	G	H
1	E.ON_Date	E.ON_60m RSD Wind DIR	E.ON_60m Wind Speed	E.ON_60m Wind Ve	E.ON_Quality (Station Height	E.ON_60m Wind Turbul	E.ON_Turbu. Quality (Station Height	E.ON_Timestamp
2	5/6/2014 20:00	177.9	10.03	0.26	100	0.14	100	5/6/2014 13:50
3	5/6/2014 20:50	167.2	10.65	-0.21	99	0.11	99	5/6/2014 14:40
4	5/7/2014 8:00	166.1	6.71	-0.03	99	0.04	99	5/7/2014 1:50
5	5/7/2014 17:50	166.4	10.05	-0.05	99	0.09	99	5/7/2014 11:40
6	5/7/2014 19:30	168.8	10.91	0.09	99	0.09	99	5/7/2014 13:20
7	5/8/2014 6:00	168.5	8.11	-0.09	99	0.09	99	5/7/2014 23:50
8	5/8/2014 6:10	168.7	8.85	-0.2	99	0.1	99	5/8/2014 0:00
9	5/8/2014 6:20	173.2	8.08	-0.01	100	0.09	100	5/8/2014 0:10
10	5/8/2014 6:30	170.7	7.21	-0.02	99	0.1	99	5/8/2014 0:20
11	5/8/2014 6:40	166.7	8.15	-0.16	99	0.07	99	5/8/2014 0:30
12	5/8/2014 6:50	174.8	8.3	0.1	99	0.09	99	5/8/2014 0:40
13	5/8/2014 7:00	172.6	8.63	0	100	0.1	100	5/8/2014 0:50
14	5/8/2014 7:10	174.9	9.48	-0.04	99	0.08	99	5/8/2014 1:00
15	5/8/2014 7:20	174.4	8.77	-0.06	99	0.09	99	5/8/2014 1:10
16	5/8/2014 7:30	166	9	-0.13	99	0.08	99	5/8/2014 1:20
17	5/8/2014 7:40	170.4	8.51	-0.12	99	0.08	99	5/8/2014 1:30
18	5/8/2014 7:50	166.4	6.59	-0.03	100	0.12	100	5/8/2014 1:40
19	5/8/2014 8:00	174.3	8.04	0.07	99	0.09	99	5/8/2014 1:50
20	5/9/2014 0:10	170.5	8.23	-0.02	99	0.1	99	5/8/2014 18:00

Step 2: Prepare your dataset: Triton Data

For Triton data, you will need to download the corrected data and run that through the tool.

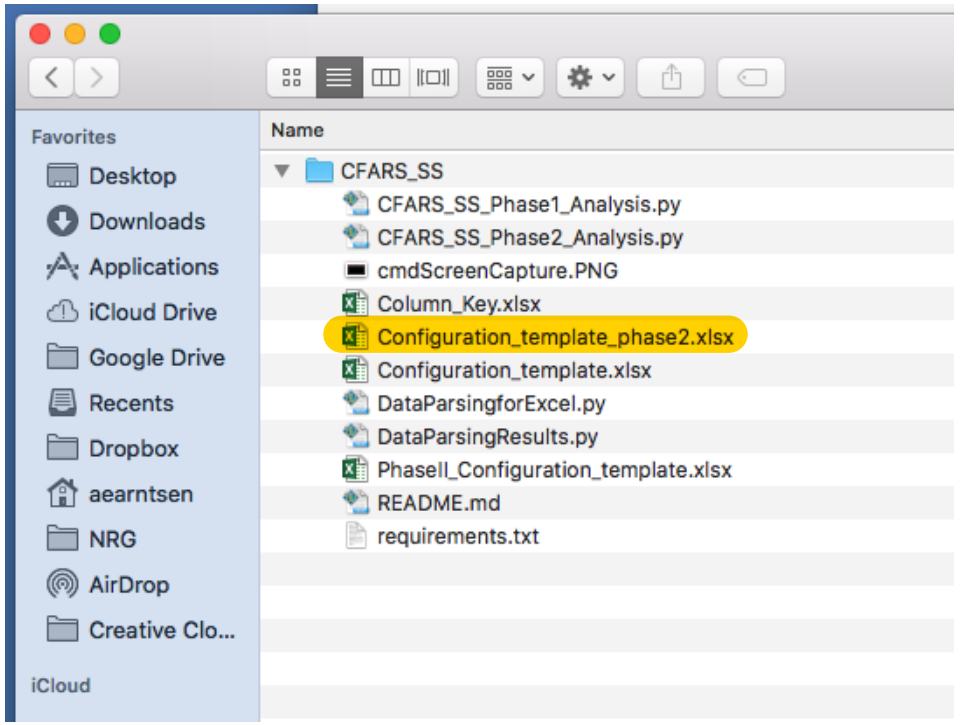
- Login to ‘https://energy.vaisala.com<https://energy.vaisala.com/>’
- Navigate to the Triton you want to correct
- Set the start and end dates
- Download the data. This data will be identical to the data you downloaded and used for Phase I, except for the addition of the new, machine-learning Turbulence Intensity measurement, with column names, for example, ‘100m TI’
- Apply QF>90 and |vert| < 1.5 m/s, as usual
- Apply any other filters that were applied to the Phase I Triton data
- Include the phase 1 triton data (pre-correction) in the data file as well

- for assistance contact **andrew.hastingsblack@vaisala.com**

Step 3: update the “Configuration template” file to match your dataset

- The “Configuration template” is a master code related file that is automatically pushed from the open source platform to the CFARS_SS on your computer (once you install required software and clone repository as noted in earlier steps). It instructs the master code on how to connect your dataset’s variable names to the master code’s variable names. The name for the phase 2 configuration is: **Configuration_template_phase2.xlsx**

- Remember, the master code provides TI bias stats one project at a time. So if you have 5 project datasets, you will set up the “Configuration template” for each project dataset, run the master code on each project dataset, and in the end have 5 result matrices to submit



Step 3: update the “Configuration template” to match your dataset

- Open the “Configuration template” and begin matching your project dataset’s variables names to the master code’s variable names. You can rename this file for each project as you’d like.
    - “Header_YourData” = your dataset’s variable names. Copy and paste your one row header information from your project dataset and *transpose* paste your variable names under in the “Header_YourData” column.
    - “Header_CFARS_Python” = the master code’s variable names. Note you can change the variable by navigating through all possible variable names from the drop down option in each cell. Note: scroll down to see all variable options.
    - Note: You do not need to match every “Header_YourData” variable with “Header_CFARS_Python” variable, only match variables that exist in your dataset (skip the others). The minimum number of CFARS_Column variable assignment is seven (RSD_WS, RSD_TI, RSD_SD, Timestamp, Ref_WS, Ref_SD, Ref_TI). The code can handle anemometer to anemometer comparisons, just leave out the RSD.
        - Other options of CFARS_Column variables are: corrWS_RSD_TI (your RSD TI measurement with a wind speed correction applied), corrWS_RSD_WS (your RSD wind speed measurement with only wind speed correction applied), corrTI_RSD_TI (your RSD TI measurement with only a TI correction applied – in case you have different method from wind speed correction only), and finally corrTI_RSD_WS (your RSD wind speed measurement from data that had the TI correction applied).

Header_YourData	Header_CFARS_Python
Time	Timestamp
PM2_Ane_A1_Avg	Ane2_WS
PM2_Ane_A1_Min	
PM2_Ane_A1_Max	
PM2_Ane_A1_Std	Ane2_SD
TI	Ane2_TI
PM2_Ane_B1_Avg	Ref_WS
PM2_Ane_B1_Min	
PM2_Ane_B1_Max	
PM2_Ane_B1_Std	Ref_SD
TI_cup	Ref_TI
TI_Lidar	RSD_TI
WS93_Avg	RSD_WS
tiE_35	
tiE_55	
40m Wind Direction	
40m Wind Speed	
WS93_Std	RSD_SD
Quality (Station Height 40m)	
60m Wind Direction	

Note: RSD_SD and Ref_SD are required to implement correction methods

Remember, the master code provides TI bias stats one project at a time. So if you have 5 project datasets, you will set up the “Configuration template” for each project dataset, run the master code on each project dataset, and in the end have 5 result matrices to submit

Step 3: update the “Configuration template” to match your dataset

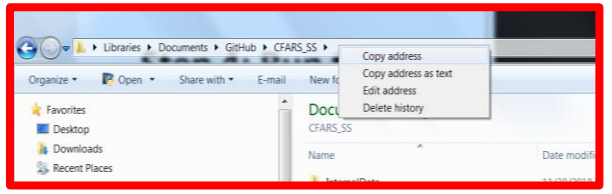
In addition to providing column matching information, the configuration template for phase 2 also asks for some site meta data and asks for info on corrections that have been applied to the data.

Site Metadata	Selection	If Option Selection is Other, Note Here
CFARS Region:	10 - North America	
Terrain:	Complex	
Comparison Height (m):	30	
Season:	Spring (Mar, Apr, May)	
Distance between RSD and Mast:		
RSD Type:	WindCube v2	
Reference Anemometer Type:	NRG #40	
Anemometer 2 Type:	Other	
Reference Anemometer Class:	1	
Anemometer 2 Class:	1	
Does anemometer mounting meet Measnet/IEC Standards?		

Key Cleaning Flags/Filters Applied	Selection	Notes
Tower Distortion using Conservative Method?	Y	
Anemometer Icing?	Y	
Wind Speed Outliers?	Y (please note thresholds)	
Sensor Degradation?	Y	
RSD Low Data Availability Filter?	Y	
RSD Low CNR Filter?	Y	
Application of site-specific transfer function (slope/offset)	N	

Step 4: Run the code for a *single* dataset

- Navigate to your start menu and search for “cmd” which will open your Windows command window
- Now indicate in the command window for your computer to run the python code. To run the code you must point to specific files in the following order: input file (i.e., python code), configuration file, output file
  - *Note windows can only read “qualified” file names, meaning you must directly tell the computer the address to find the files (see screenshot below):
  - Below is a step-by-step example how to point to your input, configuration, and output files.
    - First: Change directory (“cd”) and point to your CFARS_SS folder (the press enter and “C:\Users\a47499\Documents\Github\CFARS_SS>” will appear)
      - Ex: Type “cd C:\Users\a47499\Documents\Github\CFARS_SS“
    - Second: Point to the python code (then press space)
      - Ex: Type “python CFARS_SS_Phase2_Analysis.py”
    - Third: Point to your project dataset (then press space)
      - Ex: Type “C:\Users\a47499\Documents\Github\CFARS_SS\InternalData\EON_Project1.xlsx”
    - Fourth: Point to your configuration template (then press space)
      - Ex: Type “C:\Users\a47499\Documents\Github\CFARS_SS\Configuratin_template_phase2.xlsx”
    - Fifth: Point to where you want the output saved (i.e., where you want your result matrix to be saved and what you want to name it (finally press enter)
      - *Please save your project output with the following naming convention: Phase2Tests_ResultsMatrix_OrganizationName_ProjectX_v0X_Date
      - Ex: Type “C:\Users\a47499\Documents\Github\CFARS_SS\Phase2Tests_ResultsMatrix_EON_Project1_v01_20181130.xlsx”



Shortcut: You can copy and paste the file’s address by right clicking on the address bar and selecting “Copy address”. Then right click in the CMD window and select “Paste”

Reminder: Close all files you refer to in command window, to avoid errors.

Example from my MAC:

```
(base) Alexandras-MacBook-Pro:CFARS_SS aearntsen$ python CFARS_SS_Phase2_Analysis.py /Users/aearntsen/CFARS_Phase2_Work/Repository/cfars-ss/InternalData/
CFARS_DummyData_v02.xlsx /Users/aearntsen/CFARS_Phase2_Work/Repository/cfars-ss/Configuration_template_phase2.xlsx /Users/aearntsen/CFARS_Phase2_Work/Pha
se2Tests_ResultsMatrix_NRG_project1_8_09.xlsx
```

Step 5: Check your project results matrix

- An example file of what your output should look like will be shared with the group.
  - Note the average mean stats and above/below rated wind speed stats were calculated using 0.5 m/s bin size.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1		m	c	rsquared	ws_diff																						
2																											
3	WS_regress	1.004966	-0.05972	0.99854	0.022962																						
4	WS_regression_corrTI_RSD_Ref																										
5	WS_regression_corrWS_RSD_Ref																										
6	WS_regress	0.999059	-0.00421	0.999971	0.011164																						
7																											
8	TI_RSDvsRE	m	c	rsquared	rmse																						
9		0.697197	0.014707	0.869985	0.026954																						
10																											
11																											
12	Total Coun	4298																									
13																											
14		sensor	height	correction	m	c	rsquared	ws_diff																			
15																											
16	TI_regressic	WindCube	30	GE	0.705836	0.014221	0.850145	0.012011																			
17																											
18																											
19		RSD_WS																									
20		count																									
21		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17										
22																											
23	RSD_WS	55	201	342	483	603	707	532	372	306	294	220	107	46	17	11	2										
24																											
25		RSD_WS																									
26		count																									
27		0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	
28																											
29	RSD_WS	0	0	0	28	60	97	155	166	205	237	276	305	324	381	298	247	263	176	147	166	151	145	128	119	79	
30																											
31		TI_error_RSD_Ref																									
32		mean																std									
33		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	2	3	4	5	6	7	8	9	10	
34																											
35	TI_error_RS	0.530752	0.371653	0.281378	0.268481	0.162889	0.205395	0.072673	0.032242	0.059469	0.067725	0.068657	0.077179	0.0859	0.039699	0.052439	0.028607	0.592365	0.469282	0.465222	0.506423	0.40005	0.743339	0.315105	0.152667	0.138494	0.116
36																											
37																											
38		TI_error_RSD_Ref																									

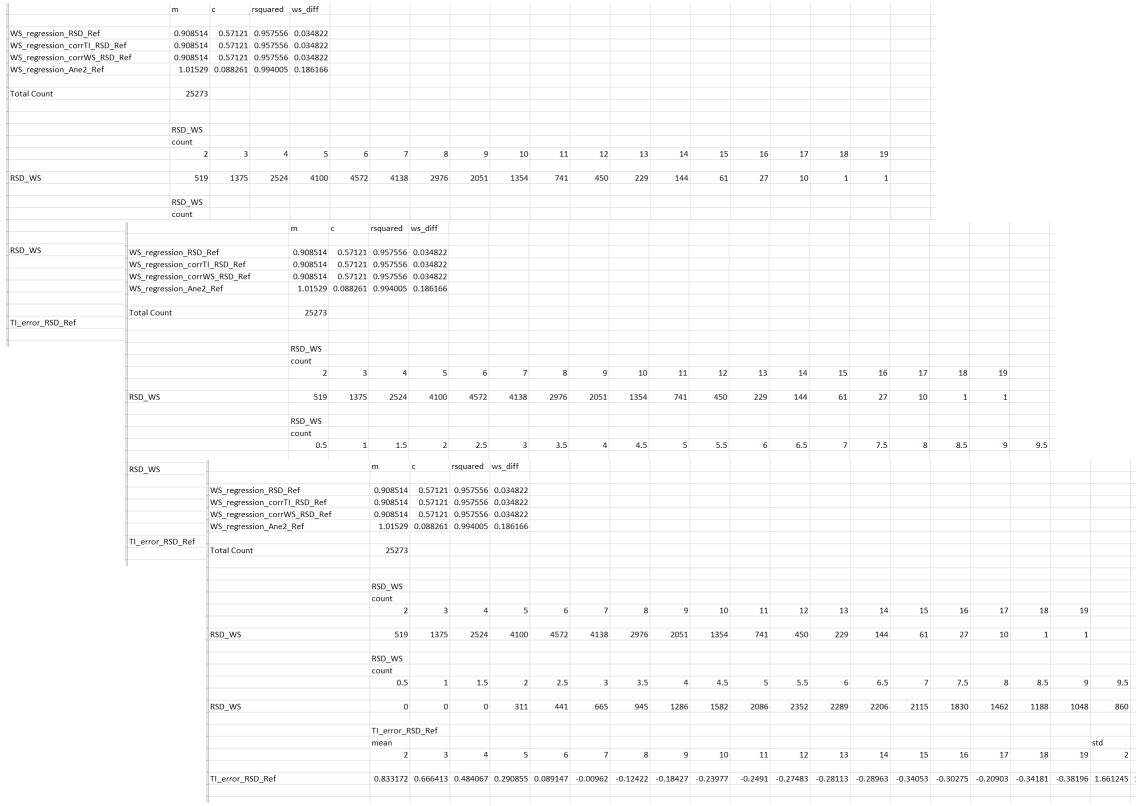
Step 6: Repeat Steps 3-5 for each project dataset

- Remember, the master code provides TI bias stats one project at a time. So if you have 5 project datasets, you will set up the "Configuration template" for each project dataset, run the master code on each project dataset, and in the end have 5 result matrices to submit

Step 10: Submit your result matrices

Email your Phase 2 analysis individual project result matrices (i.e., unaltered code output) titled "CFARS_SS_Subgroup_Phase1Tests_Results Matrix_OrganizationName_v0x_Date" to:

- AAlexandra.st.-pe@eon.com
- aea@nrgsystems.com



	m	c	rsquared	ws_diff
WS_regression_RSD_Ref	0.908514	0.57121	0.957556	0.034822
WS_regression_corrTI_RSD_Ref	0.908514	0.57121	0.957556	0.034822
WS_regression_corrWS_RSD_Ref	0.908514	0.57121	0.957556	0.034822
WS_regression_Ane2_Ref	1.01529	0.088261	0.994005	0.186166

Total Count	25273

	RSD_WS																		
	count																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
RSD_WS	519	1375	2524	4100	4572	4138	2976	2051	1354	741	450	229	144	61	27	10	1	1	

	RSD_WS																										
	count																										
	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5								
RSD_WS	0	0	0	311	441	665	945	1286	1582	2086	2352	2289	2206	2115	1830	1462	1188	1048	860								

	TI_error_RSD_Ref																										
	mean																										std
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	2								
TI_error_RSD_Ref	0.833172	0.666413	0.484067	0.290855	0.089147	-0.00962	-0.12422	-0.18427	-0.23977	-0.2491	-0.27483	-0.28113	-0.28963	-0.34053	-0.30275	-0.20903	-0.34181	-0.38196	1.661245								

You successfully processed your data!

If not, and you need support, contact:

Alexandra Arntsen: aea@nrgsystems.com