Big Data Analysis using MATLAB: Connections Between Atmospheric Muon Count and Local Weather Patterns in Atlanta, GA

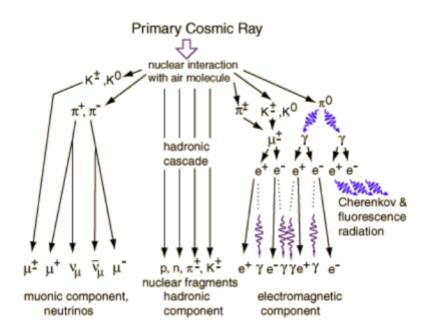
Lucius Anderson, Georgia State University Departments of Mathematics and Statistics, Computer Science, and Physics April 2017

Sections:

- Introduction
- Data Accumulation
 - Where Data comes from, and dates
 - Factors to be analyzed
- Data processing
 - How Tables where averaged
 - Tables
- Factor Studies/Analysis/Plots
 - o Factors in Response to Muon Count
 - Temperature
 - Wind Speed
 - Pressure
 - Humidity
 - Dew Point
 - Cloud Layer Coverage
 - Humidity vs Dew Point Analysis
 - Multivariable Analysis
 - Muon Count vs Humidity and Temperature
- Conclusion
- Accreditation/ Acknowledgements
 - Bibliography

Introduction

Every day in Earth's atmosphere is bombarded by thousands of Cosmic Rays. With these interactions, subatomic particles of all types are given off, many with different properties. The area of focus for our project is the muon sub atomic particle. The muon is especially important because of its qualities. It is vastly like the electron with it being a negatively charged but unlike the electron the muon has a mass that is 207 times heavier than the electron. Because of this the muon is a very fascinating particle to study but what exactly affects muon count? After much thought our group decided that a study of local weather factors would be great to see thse local weather conditions affect muon count, and if they do how so.



(Figure 1: Depicts Cosmic Ray Activity and breakdown of subatomic particles released after nuclear interaction with air molecules)

Data Accumulation

After preliminary research and introducing of the physics facilities by Dr. He, the time came to accumulate data for the project for processing, and analysis. Data accumulation was not initially difficult because of the sheer amount of different data sets that were widely available, but after looking through many sites some data sets were a lot more tedious to deal with than others. Problems such as unorthodox file types, sets with no labels, and some data set files were just too large. However, we did find a suitable website where we could find data sets to analyze: **MesoWest.** Thus, project data comes two viable locations: Dr. He directly for the Muon Particle count and MesoWest database for local weather factor data accumulation.

MesoWest is an ongoing cooperative project, started in 1996, to provide access to current and archive weather observations across the United States. Data are collected from a variety of organizations. Some stations participate in voluntary weather observing networks such as the Citizen Weather Observer Program. Others are part of mesonets that are managed by private firms or federal/state/local agencies. These data are available for a multitude of uses. Over 20,000 weather stations actively report to the MesoWest database.

Once the locations of extracting the data were finalized, the factor analysis was then narrowed six factors: cloud layer of 1, temperature, wind speed, dew point, humidity and atmospheric pressure. A set period of sixty-nine was also established because at the time Dr. He's muon data set started January 01, 2017 and ended March 17, 2017; a total of sixty-nine days.

Dr. He's data sets are accumulated by a total of four devices: A Pot Detector, Four-Paddle Detector, Double-Paddle Detector, and Weather Station. All devices except the Weather Station detector are located on the 9th floor of Langdale hall. The Weather Station device is installed on the roof of the Natural Science Center at Georgia State University, which is also beneficial for Dr. He's research because of elevation difference.



(Figure 2: Double-Paddle Detector captured from my cellular device)



(Figure 3: Another angle of the Double-Paddle Detector captured from my cellular device)

Data Processing

With the data sets compiled it became time to process the data and figure out how we would average results for analysis. Many factors went it to how this process would work; time, degree of error, and function creation. Initially we did try to process data points in Matlab using defined Matlab functions and even user created functions, however because of the way the data is structured, generation of averaging functions would require either different programming languages or averaging in Microsoft excel. In the end, we decided to average and process values for direct importation in Matlab using Microsoft Excel. Although this was not intended from the beginning, it worked out because of Andira's knowledge with using Microsoft excel, however from Andira's perspective, averaging values provided quite time consuming.

Tables

CSV files proved to be the easiest for Matlab to handle so after processing and averaging data files a direct table was created that is used directly in our project. Even after data was compiled and arranged into one main CSV file, a creation of tables using the averaged data was still created into Matlab. Because of how linear modeling is created in Matlab a creation of data tables using Matlab functions were also created. Below are example of the process of data processing is shown using Microsoft Excel and Matlab table creation for our project.

	# STATION:			
2			Jackson Atlanta International Airport	
3	# LATITUDE: 33.64028			
4		XE: -84.42694		
5	# ELEVATION	N [ft]: 1027		
6	# STATE: GA	l.		
7	Station_ID	Date_Time	air_temp_set_1	
8			Fahrenheit	
9	KATL	12/31/2016	42.8	
10	KATL	12/31/2016	42.8	
11	KATL	12/31/2016	42.8	
12	KATL	12/31/2016	42.8	
13	KATL	12/31/2016	42.8	
14	KATL	12/31/2016	42.8	
15	KATL	12/31/2016	42.8	
16	KATL	12/31/2016	42.8	
17	KATL	12/31/2016	42.8	
18	KATL	12/31/2016	42.8	
19	KATL	12/31/2016	42.8	
20	KATL	12/31/2016	42.08	
21	KATL	12/31/2016	42.8	
22	KATL	12/31/2016	42.8	
23	KATL	12/31/2016	42.8	
24	KATL	12/31/2016	42.8	
25	KATL	12/31/2016	42.8	
26	KATL	12/31/2016	42.8	
27	KATL	12/31/2016	42.8	
28	KATL	12/31/2016	42.8	
29	KATL	12/31/2016	42.8	
30	KATL	12/31/2016	42.8	
31	KATL	12/31/2016	42.8	
32	KATL	12/31/2016	42.8	
33	KATL	12/31/2016	42.98	
34	KATL	12/31/2016	42.8	
35	KATL	12/31/2016	42.8	

(Figure 4: an example of Temperature data obtained from MesoWest. Each file had an excess of 25,000+ data points and each day was averaged.)

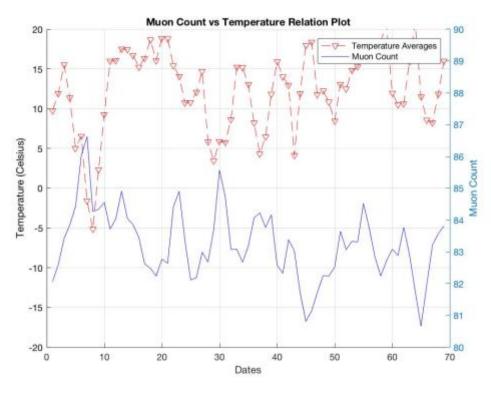
Averages								Relative H		-			vt.
		Averages		Averages		Averages		Averages		Averages		Averages	
Dury	Count	Day	Celsius	Day	Pascals	Duy	Wind Speed		Humidity	Day	Thickness	Duy	Dew Poi
	82.06662		9.776923		98320.5	1			94.90145		184.9057		7.46589
	82,59931		11.89135		98264.63		3,334844		99.66756		155.0926		12.0372
	83.41389		15.53197		944.84		3.917175		87.32997		97.86885		12.0775
	83.84861		11.41493		97444.45		5.387396		70.98463		33.33333		3.77762
	84.42708		5.837544		97709.15		3.409649		57.06864		86,14865	5	-1.3836
	86.02986		6.538811		97654.87		4.259336		81.32104		148.1073	- 6	
7	86.67292	7	-1.54757	7	98097.64	7	6.891007	7	73.37901	7	65.09434	7	-8.5017
8	84.26944	8	-5.1316	8	99375.92	8	6.640833	8	53.18737	8	7.457627	8	-12.1
9	84.34097	9	2.344964		99726.43	9	2.986751	9	38.05528	9	27.73723	5	-11.695
10	84.55556	10	9.271731	30	99055.76	10	4.266714	10	61.75489	10	78.61842	10	2.8367
11	83.71597	11	16.00903	11	98786.06	11	3.915729	11	65.61043	11	96.60194	11	8.72061
12	84.04236	12	16.04097	13	98972.07	12	2.872535	12	76,7485	12	50.96463	12	11.734
13	84.91042	1.3	17.54653	13	99338.54	13	1.823611	13	80.7923	13	88.94231	13	13.6274
14	84.05278	14	17.44236	14	99314.75	14	1.171667	14	70.70709	14	31.43322	14	11.4599
15	83.85278	15	16.69688	15	98923.58	15	1.429826	15	61.40369	15	22.59136	15	4.7990
16	83.45347	16	15.24479	16	98735.76	16	1.30934	16	78.58571	16	85.01629	16	12.5321
17	82,62569	17	16.37222	17	98602.88	17	3.220968	17	75.68129	17	75.16234	17	14.1912
18	82.48506	18	18.68715	11	98326.81	18	4.189271	18	72.12018	18	66.28289	11	11.4933
	82,23958	19	16,05035		98125.51	19	2.462882		76,39009	19	79.56811	15	14.1080
	82,77222	20			97508.34		3.796551		89,43638		85,24845		15,0853
	82.64028		18.82639		97406.16		3.554444		95.39273		128.3333		13.856
	84.41875		15.40486		96876.45		3.37/674		90.97724	22			9.9401
	84.90764		14,08438		95947.87		5,030104		71,23308		133.3333		5,40443
	83,35694		10.76181		95794.37		6.588958		54.95149		78.26797		2.52005
	82.11736		10.80804		97181.94		6.648175		62.57505		6.560294		6.47720
	82.19306		12,11029		97513.09		2.573382		58,29683		45,58824	26	
	82.99167		14.71277		97342.25		5,675693		49.58356		12,54545	27	
	82.68264		5.858712		97990.23		5.129242		50.67756		6.092437	25	
	83.69792		3.448014		38046.48		4.824188		52.61587		60.78767	25	
	85.57083		5.896585		97777.94		4.765343		48.35053		5.123675	36	
	84.73333		5.753676	31			5,596434	31	62.2169		5.731225		7.22880
	83.07778		8.659191	32	20000000		4.304632		76.88663		36.45833		8.49352
	83.08611		15.22434		98056.03		4.384045		48.73925		53.69718	33	
	82.67569		15.20786		98258.35		3,806043		32.23432		84.61538	34	
	83.17569		13.08029		98428.93		3.695448		88.63065		20,35088	35	
	84.06667		8.231673		98591.49		5.272918		78.72781	36		34	
	84.22986		4.312411	37			4.105142	37		37			13.1252
38	83.76736	38	6.466906	-	98561.77		2.524209	_	87.75606	38	92,40506	34	12.334
39		36	11.8581	25	98427.13	39	2.153768	39	49.86000	29	129.7428	35	-5.476
	82.58681	40	15.92562	40	97905.45		3.539004		41.90678	40	56.73401	40	
41	82.32361	41	14.03548	41	97267.39	41	2.480251	41	79.12696	41	5,06993	41	14.7331
42	83.38194	42	12.98175	42	97406.25	42	6.70786	42	64.84715	42	129.2388	42	8.34383
43	83.83264	40	4.124653	43	98863.09	43	5.275104	43	31.25534	43	139.5639	43	-5.6704
44	81.77292		11.91285	44	98768.05	44	3.482535	44	55.13699	44	3.636364	44	3.06963
45	80.81597	45	17.99688	45	98366.35	45	3.905694	45	41.81773	45	37.96296	45	-5.0209
46	81.14167	46	18.38576	46	98167.19	46	5.515903	46	34.29825	46	60.28037	46	-3.7897
47	81.73125	40	11.8	40	98277.87	47	4.156562	47	59.93022	47	1.724138	47	4.49185
48	82.25268	48	12.24755	48	97504.63	48	1.470944	48	75.40242	48	5.485232	48	10.163
49	82.24306	49	10.83646	45	97032.56	49	6.645208	49	58.34672	49	90.64626	45	7,4359
50	82.51806	30	8.472222	36	97874.7	50	4.731319	50	77.93929	50	69.62025	36	11.8834
51	83.64653	51	13.10139	51	97949.24	51	3.324306	51	88.67387	51	9.927798	51	15.216
52	83.07222	50	12,51014	50	97789.36	52	1.660175	52	71.99598	52	78.52564	52	12.1820
53	83.33333	58	14.80347	58	98002.25	53	3.150625	53	65.85736	53	121.3846	58	12.327
54			15.27083	54	98360	54	2.520903		41.92567		40.13841	54	
55	84.52014	56	16,63368	56	98234.82	55	3.363785		33.06451	56		56	
	83.75347		16.63785		97651.99		3.692535		71.93347		31.67203		5.2879
	82.84828		16.34549	57			4.524931		84.77482		2.173913		14.670
	82.24236		18.45417		97441.05		2.183519		63.08356		47.99331		10.30%
	82.71667		20.63299		97262.74		3.846528		40.89982		77.33119		4.822
	83.08333		11.97222		97735.81		6.898958		31.3234		99.37888		7.991
	83,76944		10.5309		98753.89		2.496399		35.61583 47.43991		12.81588		4.0235
			10.61076				3.628507				1.896552		
	82.91806		15.96528		98592.8		3.124097		60.60077		3.846154		8.8725
64			26.72188		97882.85		5.887361		69.77825		42.75618		7.8630
	80.66135		11.525		98349.01		7.243507		33.39518		119.3811		-2.2220
	81.96042		8.575694		99255.62		4.583542		57.29058		88.62179		7.33745
	83.22153		8.210764		99659.24		2.665575		40.55511		45,68966		-3.3906
620	83.56944		16,00035	61	39444.46	68	1.494792	68	87.2265	68	3.696498		1.8883

(Figure 5: Finished Table after Data Processing and Averaging)

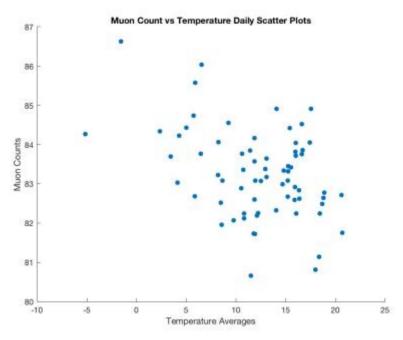
Factor Studies/Analysis/Plots

Muon Count vs Temperature

Although Dr. He's initial studies had already determined that the relationship between Muon count and Temperature was anti-correlative, we still decided that analyzing local temperature in relation to muon count would be a great area of focus. After processing the data different collections of plots were created using Matlab technology and a linear regression model was also created. However due to low R-Squared values in the model, we could only low confidence conclusions between muon count and temperature. Even though the statistical model may not be within desirable confidence bounds, correlations are still shown in the plots below.



(Figure 6: Muon Count versus Temperature Plot Over Time total of 69 Days)



(Figure 7: Scatter Plot that shows relationship between Muon Count and Temperature in Atlanta Georgia)

MuonCount_vs_TemperatureReadingLinearModel =

Linear regression model:

MuonCount ~ 1 + Temperatures

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept) Temperatures	84.351 -0.089484	0.31956 0.024081	263.96 -3.7159	8.0148e-103 0.00041512

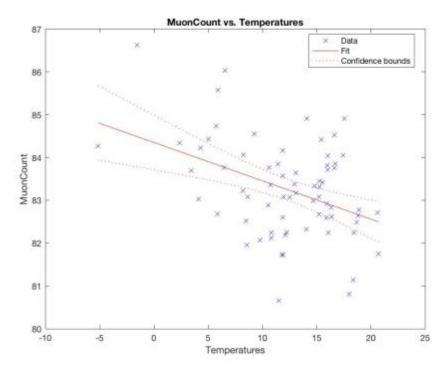
Number of observations: 69, Error degrees of freedom: 67

Root Mean Squared Error: 1.03

R-squared: 0.171, Adjusted R-Squared 0.159

F-statistic vs. constant model: 13.8, p-value = 0.000415

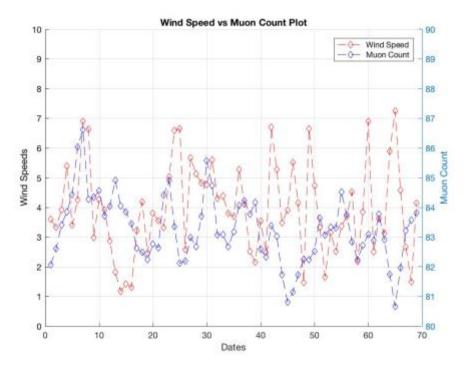
(Figure 8: Linear Regression Model of Muon Count in Response to Temperature)



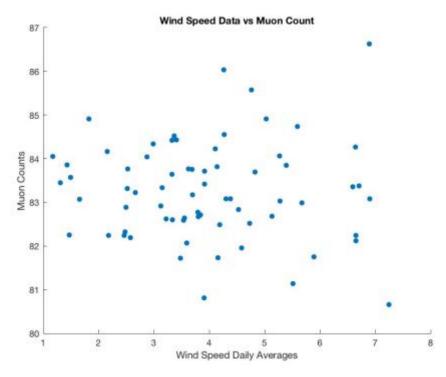
(Figure 9: Linear Regression Model of Muon Count and Temperature, with 95% Confidence bounds generated by Matlab, which supports anti-correlative relationship)

Muon Count vs Wind Speed

Wind Speed was a factor mentioned by Dr. He to analyze and after further analysis of it, wind speed proved to be a non-correlative factor. Because of low confidence bounds again on the statistical model we cannot be highly confident of our conclusion, but I still feel intuitively that wind speed does not have much effect on muon count in the atmosphere.



(Figure 10: Muon Count vs Wind Speed Values plotted over a period of 69 Days)



(Figure 11: Muon Count versus Wind Speed Scatter Plot)

MuonCount_vs_WindSpeedLinearModel =

Linear regression model:

MuonCount ~ 1 + WindSpeedReadings

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	83.367	0.38101	218.8	2.2749e-97
WindSpeedReadings	-0.027808	0.089888	-0.30937	0.758

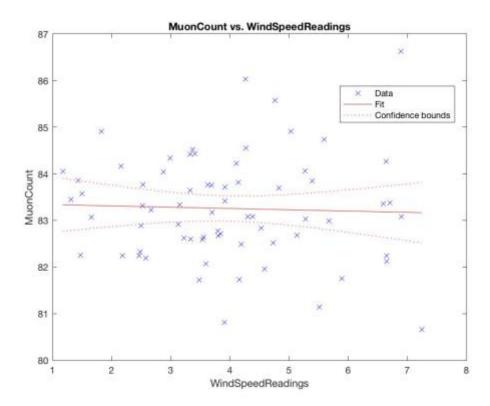
Number of observations: 69, Error degrees of freedom: 67

Root Mean Squared Error: 1.13

R-squared: 0.00143, Adjusted R-Squared -0.0135

F-statistic vs. constant model: 0.0957, p-value = 0.758

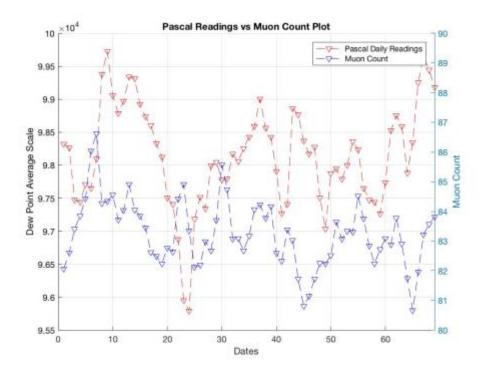
(Figure 12: Linear Regression Model of Muon Count and Wind Speed)



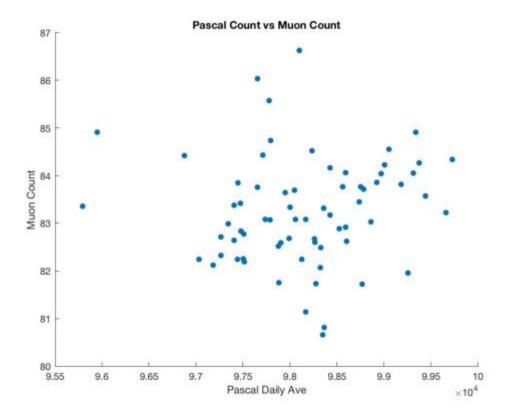
(Figure 13: Muon Count vs Wind Speed with 95% Confidence bounds generated by Matlab)

Muon Count vs Pressure

While in the preliminary stages of the project I initially thought that pressure would have a major impact on muon count just because of the fragile nature of subatomic particles. After processing and analysis my intuition proved to be wrong. Analysis shows that local atmospheric pressure has little to no effect on muon count. Pressure is relatively consistent in Atlanta so deviations in data are not high. Below are figures over this data.



(Figure 14: Muon Count vs Pressure Readings over a period of 69 Days)



(Figure 15: Muon Count vs Pressure Reading Scatter Plot)

```
MuonCount_vs_PascalLinearModel =
```

Linear regression model:

MuonCount ~ 1 + PascalReading

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	72.023	17.205	4.1862	8.4558e-05
PascalReading	0.00011446	0.0001753	0.65294	0.51603

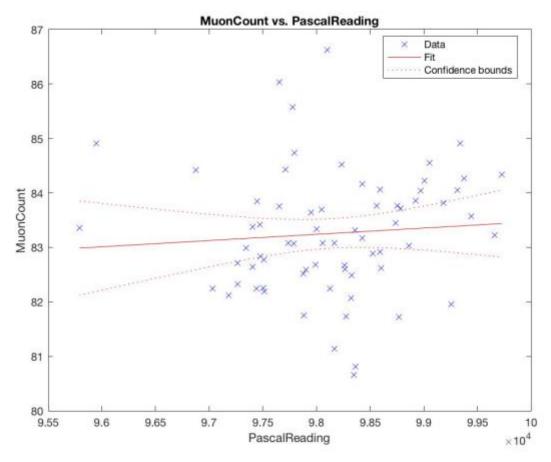
Number of observations: 69, Error degrees of freedom: 67

Root Mean Squared Error: 1.13

R-squared: 0.00632, Adjusted R-Squared -0.00851

F-statistic vs. constant model: 0.426, p-value = 0.516

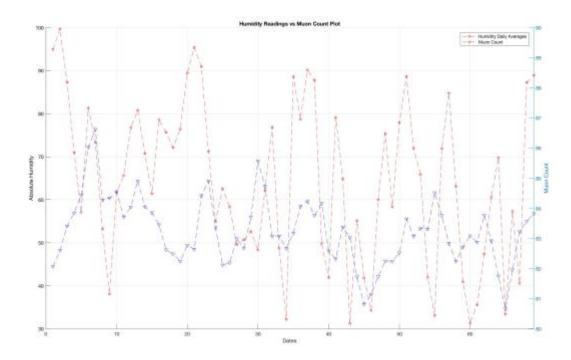
(Figure 16: Linear Model of Muon Count to Pressure, of all models this model has the worst R-squared values)



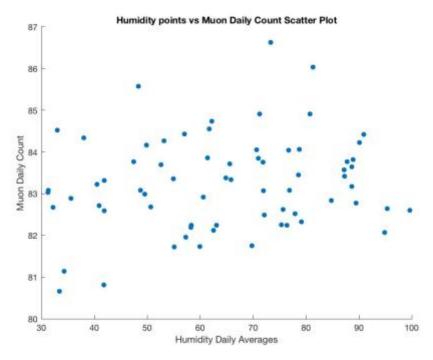
(Figure 17: Linear Model between Muon Count and Pressure Values, this figure shows a slight correlation but because of poor confidence values we cannot confirm this. With 95% Confidence intervals)

Muon Count vs Humidity

Humidity was chosen as a factor for analysis because it is an integral part of our environment. Especially living in the southeastern coast of the United States humidity places a vital role into our lives as citizens of Atlanta. Humidity proved to also be a non-correlative factor in response of muon count. R squared values also proved to be low for this model as well so confidence is not high for conclusion.



(Figure 18: Humidity vs Muon Count Plot over 69 days)



(Figure 19: Scatter Plot of Muon Count and Humidity values)

```
MuonCount_vs_HumidityLinearModel =
```

Linear regression model:
 MuonCount ~ 1 + Humidity

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept) Humidity	82.573 0.010601	0.48507 0.0072297	170.23 1.4663	4.4528e-90 0.14725

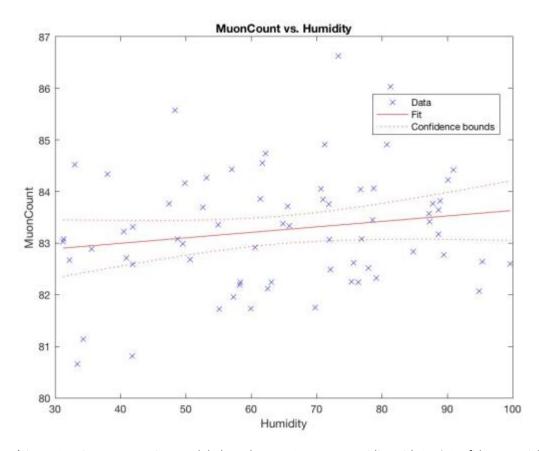
Number of observations: 69, Error degrees of freedom: 67

Root Mean Squared Error: 1.11

R-squared: 0.0311, Adjusted R-Squared 0.0166

F-statistic vs. constant model: 2.15, p-value = 0.147

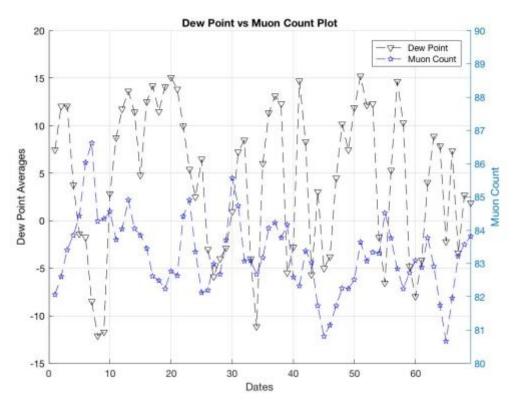
(Figure 20: Linear Model generated using Matlab function. Muon Count in Response to Humidity)



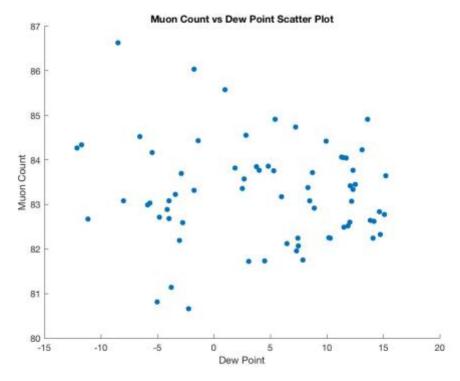
(Figure 21: Linear Regression Model Plotted Muon Count vs Humidity with 95% Confidence variables)

Muon Count vs Dew Point

Relative humidity and dew point both give us an idea of the amount of moisture in the atmosphere; however only dew point is a true measurement of the atmospheric moisture. Dew point is the term that most meteorologists use to describe the amount of moisture in the air. Because of this analysis of muon count in response to dew point seemed to be a valid point for research.



(Figure 22: Muon Count vs Dew Point Plot over a period of 69 days)



(Figure 23: Scatter Plot of Muon Count vs Dew Point, shows no correlation between two data sets)

MuonCount_vs_DewPointLinearModel =

Linear regression model: MuonCount ~ 1 + DewPoint

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept) DewPoint	83.32 -0.014656	0.15505 0.01755	537.37 -0.83511	1.6957e-123 0.40662

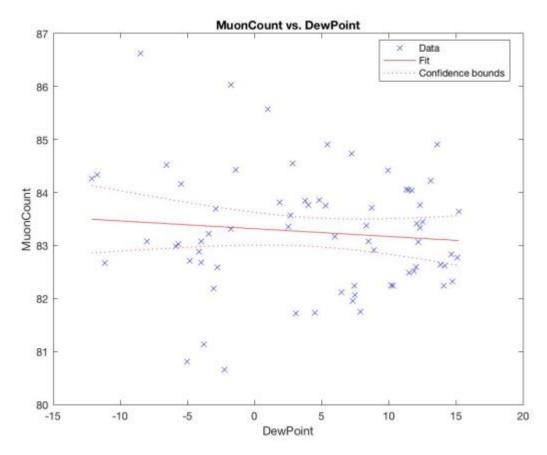
Number of observations: 69, Error degrees of freedom: 67

Root Mean Squared Error: 1.13

R-squared: 0.0103, Adjusted R-Squared -0.00447

F-statistic vs. constant model: 0.697, p-value = 0.407

(Figure 24: Linear Model for Muon Count in Response to Dew Point data)



(Figure 25: Linear Regression Model Plotted of Muon Count in response to Dew Point with 95% confidence intervals)

Cloud Layer Coverage Data

Cloud Layer data was supposed to be a pivotal factor in our project because subatomic particles are highly affected by cloud coverage. Because of this we made a huge effort to take this into account and analyze cloud coverage data. While we could accumulate the data for cloud layer coverage, the data was more qualitative than quantitative. Our data set obtained from MesoWest provided Cloud Layer coverage based on three factors: layer type, height, and cloud type. We decided to use "Cloud Layer Coverage 1" meaning that only one cloud layer was present for that day the data was taken.

The data is broken down as follows:

- Thousandth and hundreds place of the value represents height, in hundreds of feet
- The tens place digit represents the cloud type following this model
 - \circ 0 = missing
 - 1 = clear
 - 2 = scattered
 - 3 = broken
 - 4 = overcast
 - 5 = obscured
 - 6 = thin scattered
 - 7 = thin broken
 - 8 = thin overcast
 - 9 = thin obscured
- Example: 567 means a cloud is at a height of 56 = 5600 feet
- The last digit 7 identifies the cloud type as "think broken."

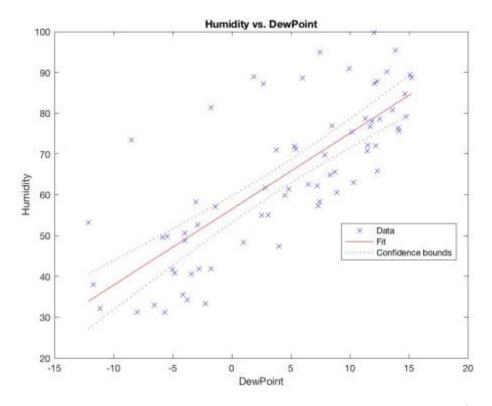
Because of this qualitative data type few conclusions could be made that would be of high relevance to the project and research. Andira did create a linear regression model between muon count and cloud layer coverage, but in my report I am choosing not to show it because it not having high relevant research value.

Humidity vs Dew Point

Because of such low confidence values on the statistical models generated in our research. A decision was made to test relations between dew point and humidity. After research, online we found out that dew point and humidity did have a positive correlation between one another, and so we tested our data and looked to see if we could come to the same conclusion. This would prove whether data processing was done correctly by our group. And after testing the two factors we received a positively correlated model between the two values and a very good R-Squared value, meaning that we could conclude that data processing for our project was done correctly.

```
DewPoint_vsHumidityLinearModel =
Linear regression model:
    Humidity ~ 1 + DewPoint
Estimated Coefficients:
                   Estimate
                                 SE
                                          tStat
                                                      pValue
    (Intercept)
                   56.49
                                1.6403
                                          34.439
                                                    2.4581e-44
    DewPoint
                    1.86
                                                      5.96e-15
                               0.18566
                                          10.018
Number of observations: 69, Error degrees of freedom: 67
Root Mean Squared Error: 11.9
R-squared: 0.6, Adjusted R-Squared 0.594
F-statistic vs. constant model: 100, p-value = 5.96e-15
```

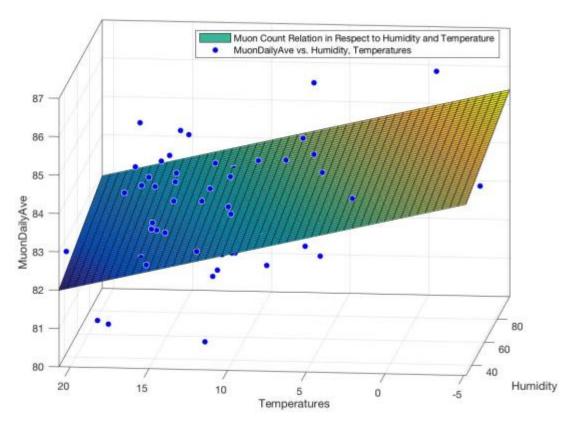
(Figure 27: Linear Model Generated through Matlab for Humidity in Response to Dew Point Data)



(Figure 28: Humidity vs Dew Point Linear Regression Model plotted using Matlab with 95% confidence bounds)

Multivariable Analysis

Also as a test of muon count in response to more than one variable at a time, we decided to plot a 3D plot of Muon Count versus Temperature and Humidity. The plot did look good, and contrary to our other factor analysis, did have a higher R squared values than other variables singularly modeled against muon count.



(Figure 29: 3D Plot of Muon Count in Response to Humidity and Temperature, most points are not however in the plane)

Conclusion

In conclusion, it would be easy to say that most factors analyzed in our project have no correlation on muon count, however due to such low confidence values we cannot conclude this thoroughly. Many plots and models such as the scatter plots and linear regression models show none correlative evidence, but very low R squared values go against this claim. It is easy to say the plot or models show no correlation but low confidence values don't provide a claim backed in good standing. Intrinsically however, I personally feel that out of all of the factors analyzed in our project, temperature and cloud layer are probably the only local weather factors that have most likely have an impact on muon count. But in the scientific community personal claims are rarely supported without scientific evidence (as it should be). For better analysis, next time I think that obtaining larger sets of data would be better, maybe obtaining data sets from conglomerate databanks such as NASA or the CDC. Averaging maybe could be done maybe in other computer programs such as python for better data processing and averaging. I believe that more research should happen for muon count in response to local weather analysis because it could be huge in areas such as cancer and medical research. Overall I enjoyed our project and I would consider further research.

Accreditation/ Acknowledgment

I would like to give many thanks to Dr. He for his research, data and enthusiasm to help us in our project, Dr. Jiang for help with the project, help with coding, and teaching the class because Matlab is an incredible tool that I am glad I learned to use. And of course my partner Andira Putri, because without her help, knowledge, and precious time our project would not be near as good as it is, and data processing and averaging would've killed me lol. Thank you

Bibliography

Citations!

Dayananda, Mathes A., "Correlation Studies of Cosmic Ray Flux and Atmospheric and Space Weather." Dissertation, Georgia State University, 2013. http://scholarworks.gsu.edu/phy astr diss/65

Horel, J.J., M.M. Splitt, L.L. Dunn, J.J. Pechmann, B.B. White, C.C. Ciliberti, S.S. Lazarus, J.J. Slemmer, D.D. Zaff, and J.J. Burks, 2002: MesoWest: Cooperative Mesonets in the Western United States. Bull. Amer. Meteor. Soc., 83, 211–225, doi: 10.1175/1520-0477(2002)083<0211:MCMITW>2.3.CO;2.

Ramesh, N., Hawron, M., "Flux Variations of Cosmic Muons." Southern Arkansas University, 2011.