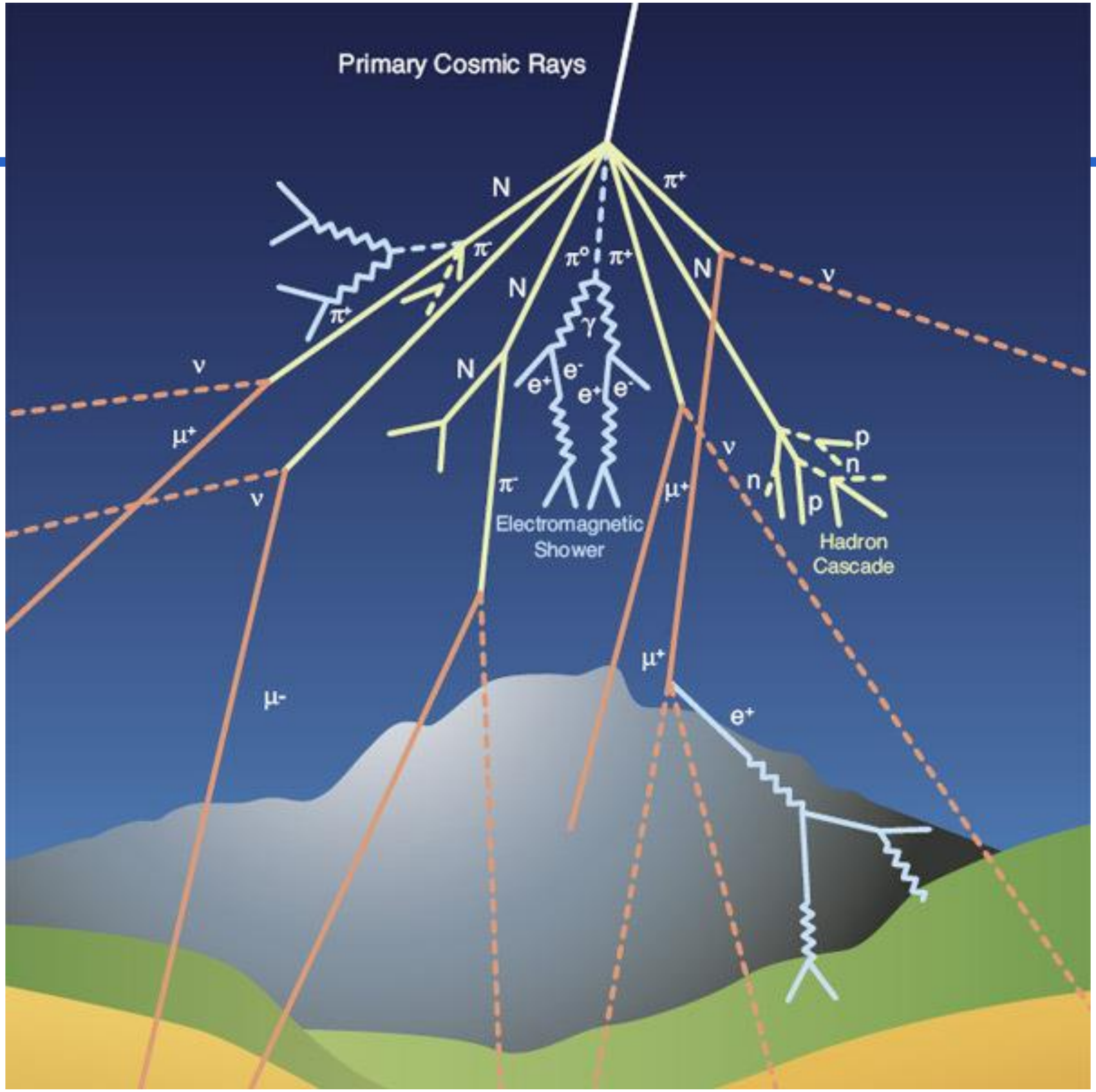
Exploring short-term correlations between space weather and muon flux using the CosmicWatch detector

Nevena Cail
Faculty mentor: Dr. Jeffery Spirko

Spring 2025 Honors Symposium
Texas A&M University-Corpus Christi

May 6, 2025

… muon?



Cosmic rays generate showers of muons in the upper atmosphere. Over 1 million muons pass through your body each minute. Variations in muon flux can be used to study the behavior of CRs in the solar system.

CosmicWatch

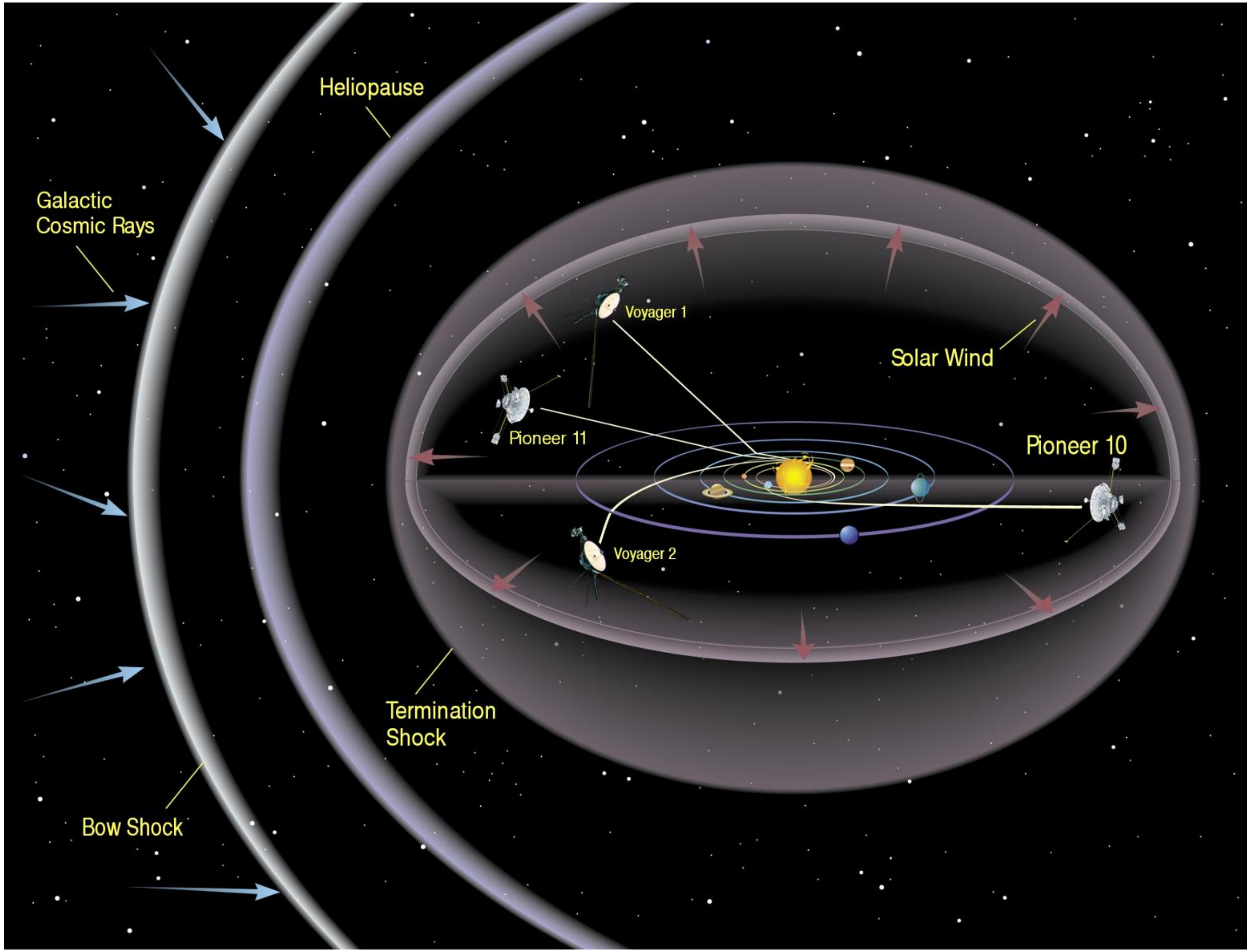
The Physics Behind the CosmicWatch Desktop Muon Detectors

December 20, 2018

The CosmicWatch Desktop Muon Detector is a Massachusetts Institute of Technology (MIT) and Polish National Center for Nuclear Research (NCBJ) based undergraduate-level physics project that incorporates various aspects of electronics-shop technical development. The detector was designed to be low-power and extremely portable, which opens up a wide range of physics for students to explore. This document describes the physics behind the Desktop Muon Detectors and explores possible measurements that can be made with the detectors. In particular, we explore various physical phenomena associated with the geomagnetic field, atmospheric conditions, cosmic ray shower composition, attenuation of particles in matter, radioactivity, and statistical properties of Poissonian process.

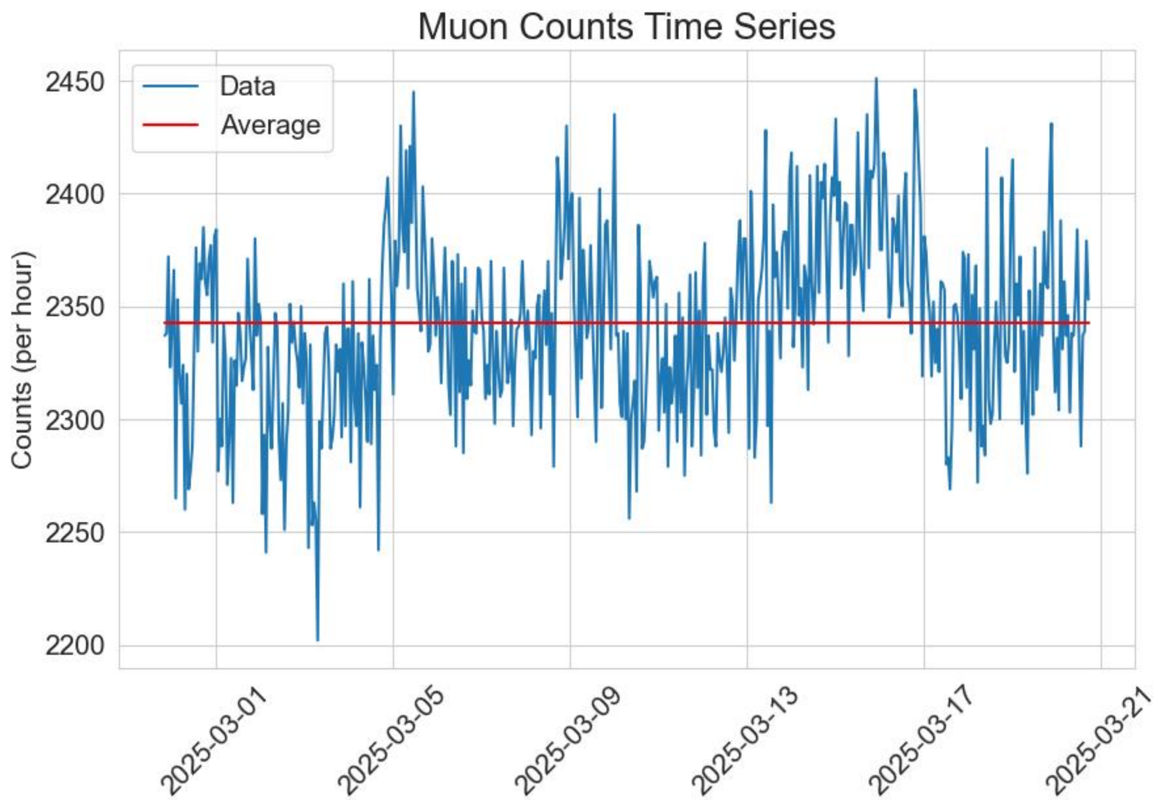
The CosmicWatch project was created at MIT so that physics undergrads could collect their own muon data. Part of this project was constructing the detector and letting it sit and record data in the TAMUCC observatory.

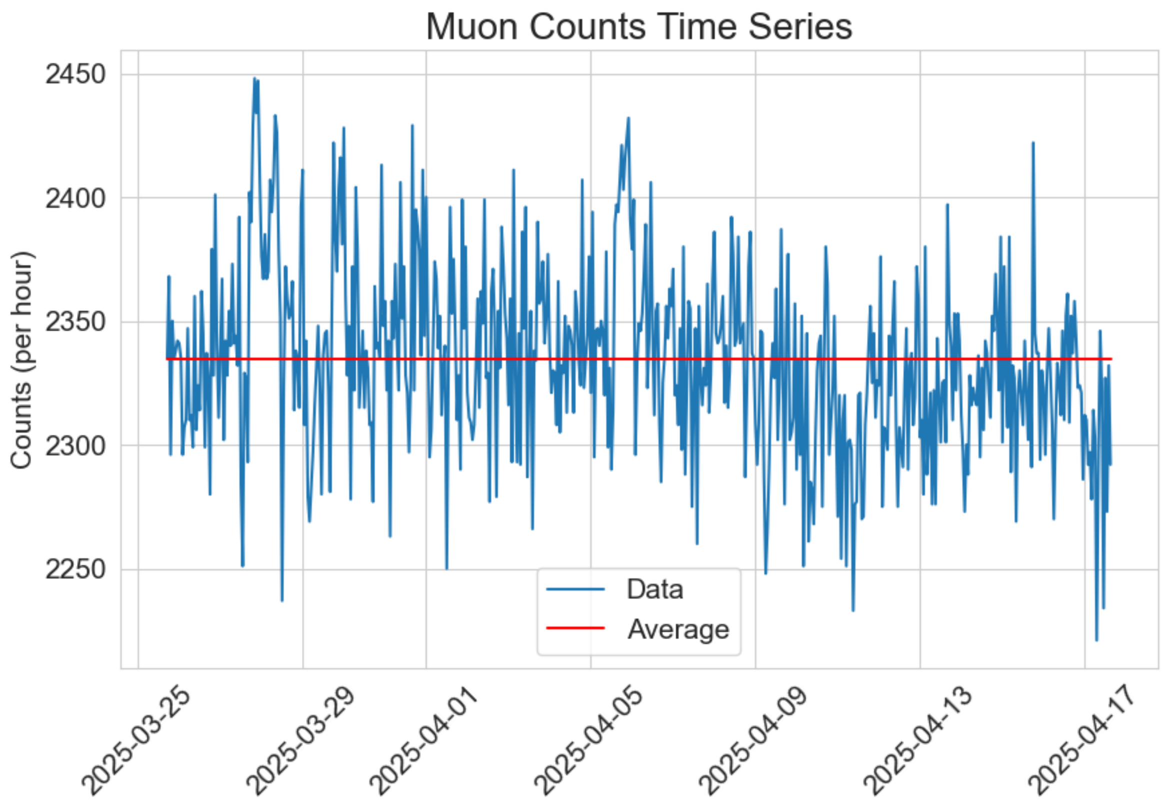
Cosmic rays



Most CRs are galactic CRs. Their trajectories through the solar system highly depend on magnetic and plasma structures. Muons can be used to study CRs if Earth's local atmospheric effects of temperature and pressure are removed. We forgo this step, and just observe patterns in our raw muon count data.

Raw data – dataset 1 & 2



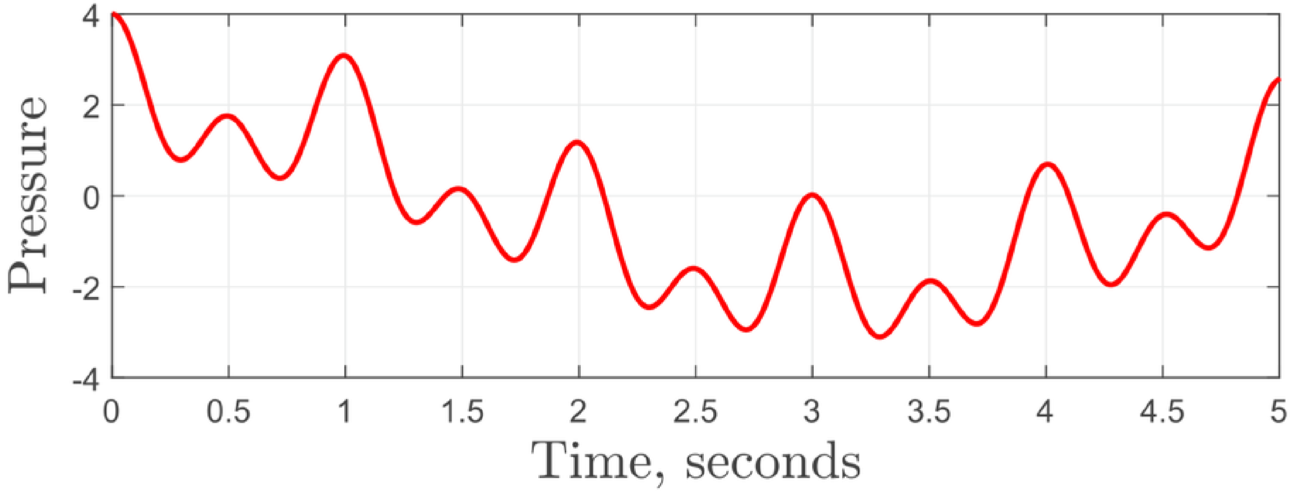


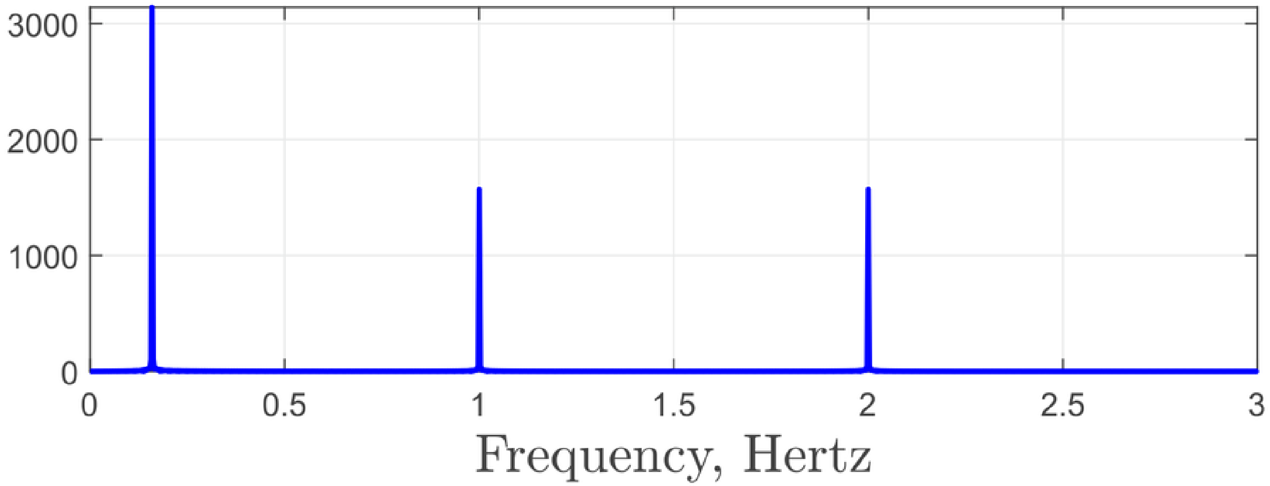
Detector counts per hour over two separate periods: 02/27 – 03/20 (21 days) & 03/25 – 04/19 (25 days).

spectral analysis

what analysis?

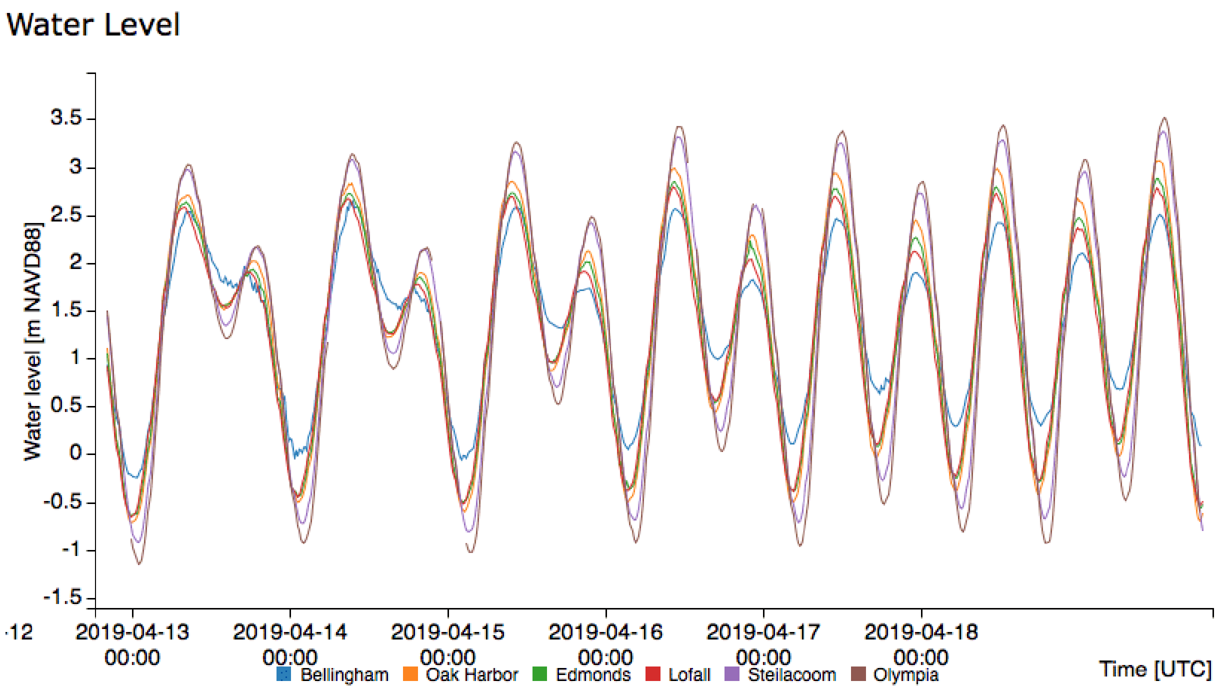
$$F(\alpha) = \int_{-\infty}^{+\infty} f(x) e^{-2\pi i \alpha x} \, dx$$





Spectral analysis is about extracting information about the frequencies in a signal. Fourier transforms do this by treating an input signal as if it is a sum of sines and cosines. It returns the magnitude of each frequency found in the signal.

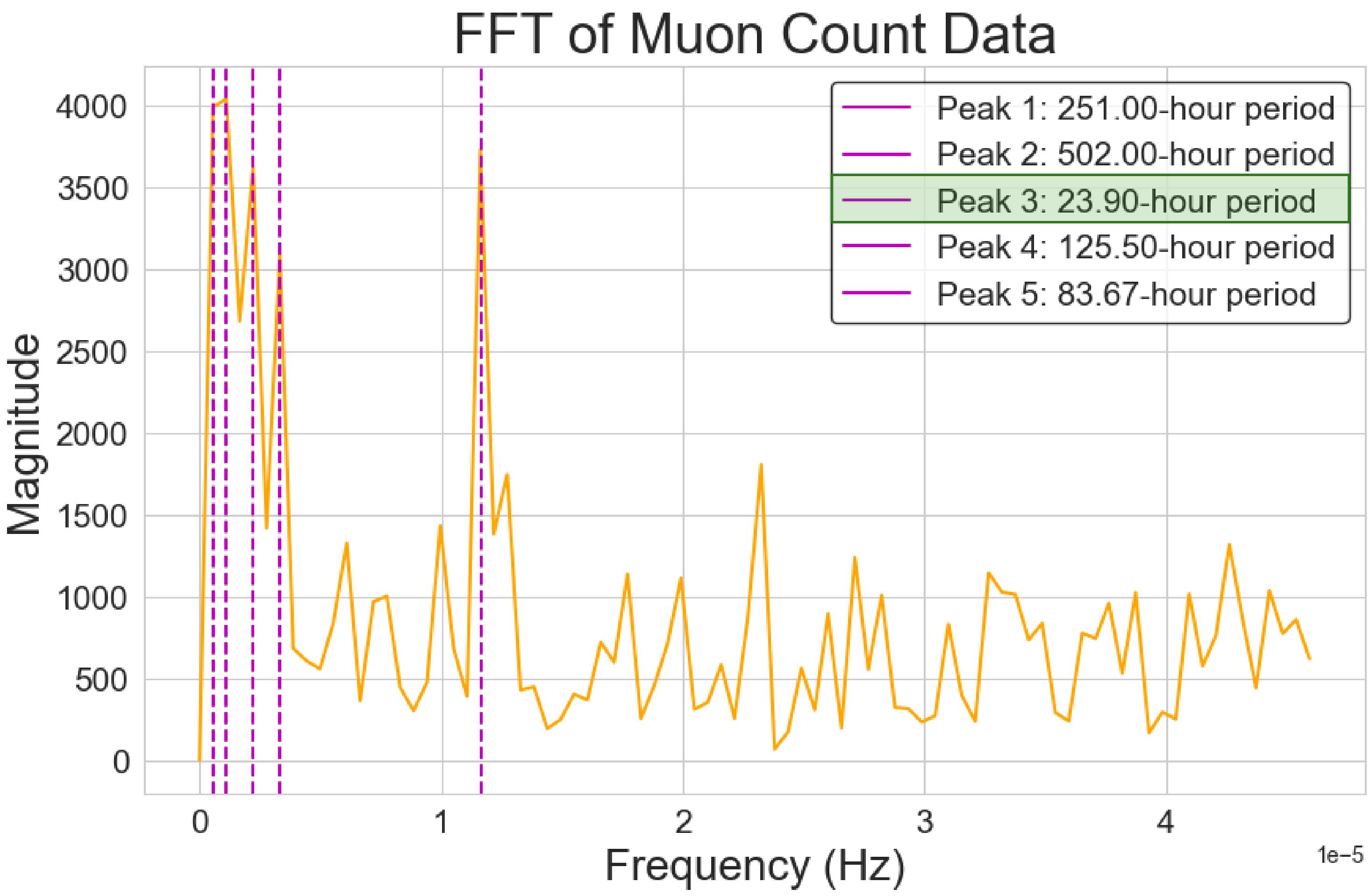
e.g. tides



Tidal constituents	Name	Equilibrium Amplitude [m]	Period [h]
Semidiurnal			
Principal lunar	M2	0.24	12.42
Principal solar	S2	0.11	12.00
Lunar elliptical	N2	0.046	12.66
Lunar-solar declinational	K2	0.031	11.97
Diurnal			
Lunar-solar declinational	K1	0.14	23.93
Principal lunar	O1	0.10	25.82
Principal solar	P1	0.047	24.07
Lunar elliptical	Q1	0.019	26.87
Long Period			
Fortnightly	Mf	0.042	327.9
Monthly	Mm	0.022	661.3
Semiannual	Ssa	0.019	4383

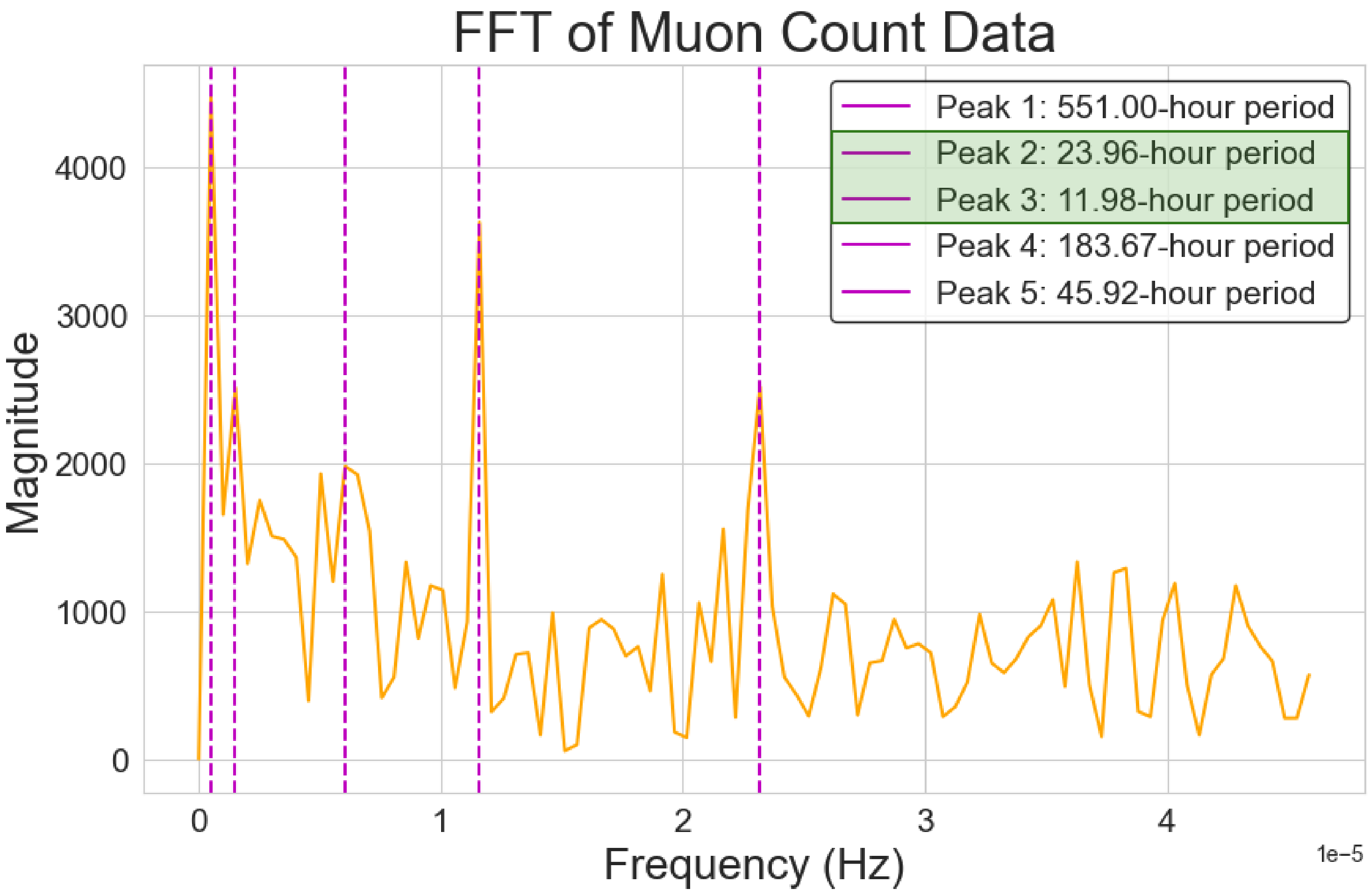
Tidal signals are made up of many “tidal constituents” on the order of hours to years due to the cyclic nature of the various gravitational forces at play. Looking at a water level time series, the frequencies that make up the signal might not be obvious, but can be found by Fourier analysis.

Inspecting the FFT (dataset 1)



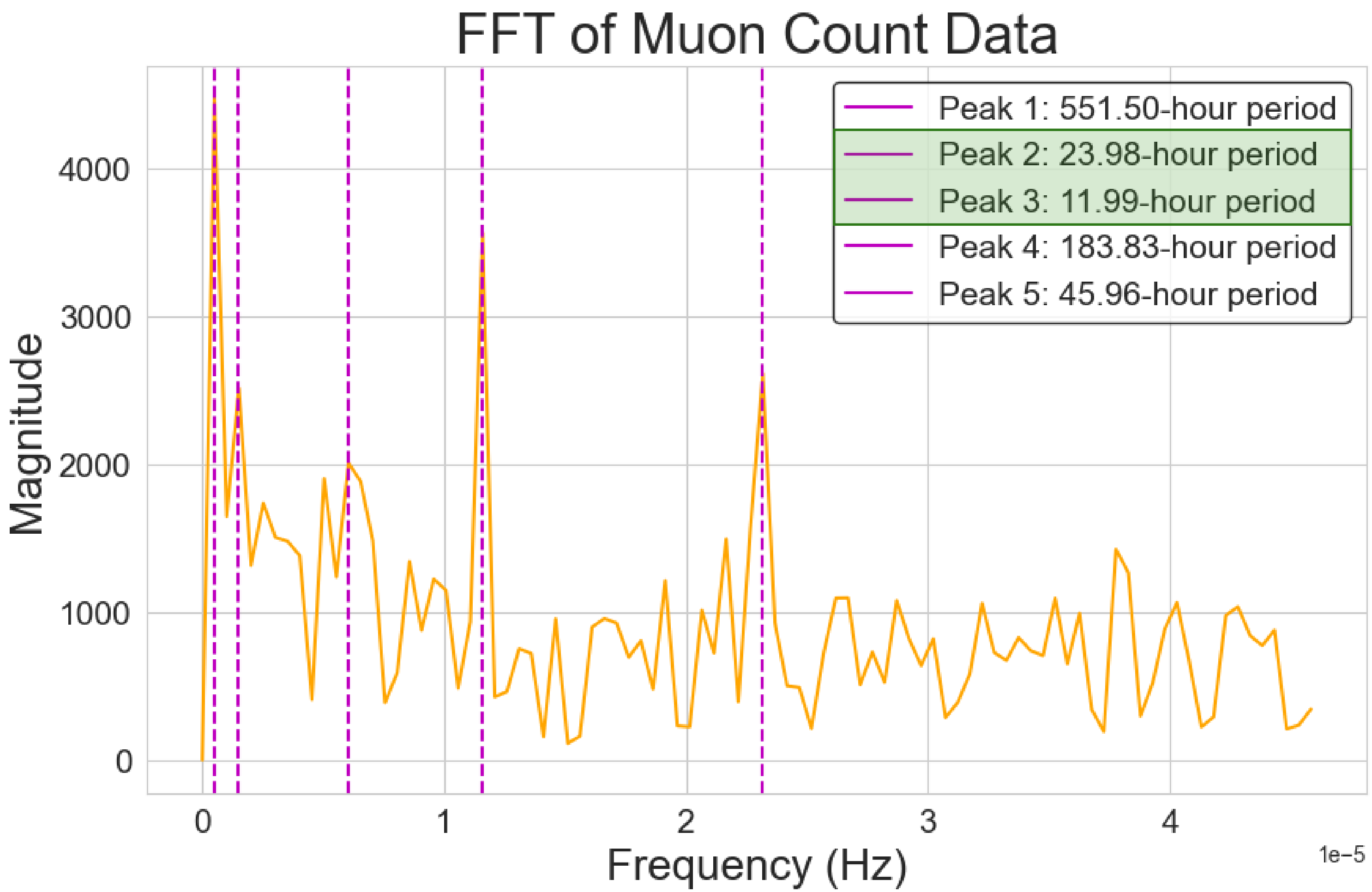
Zoomed in to exclude frequencies higher than 1/(6-hrs). Top 5 strongest peaks are listed in order of descending magnitude. Strong component corresponding to ~24 hours indicates daily temperature fluctuation might be modulating the muon flux (an expected result).

Inspecting the FFT (dataset 2)



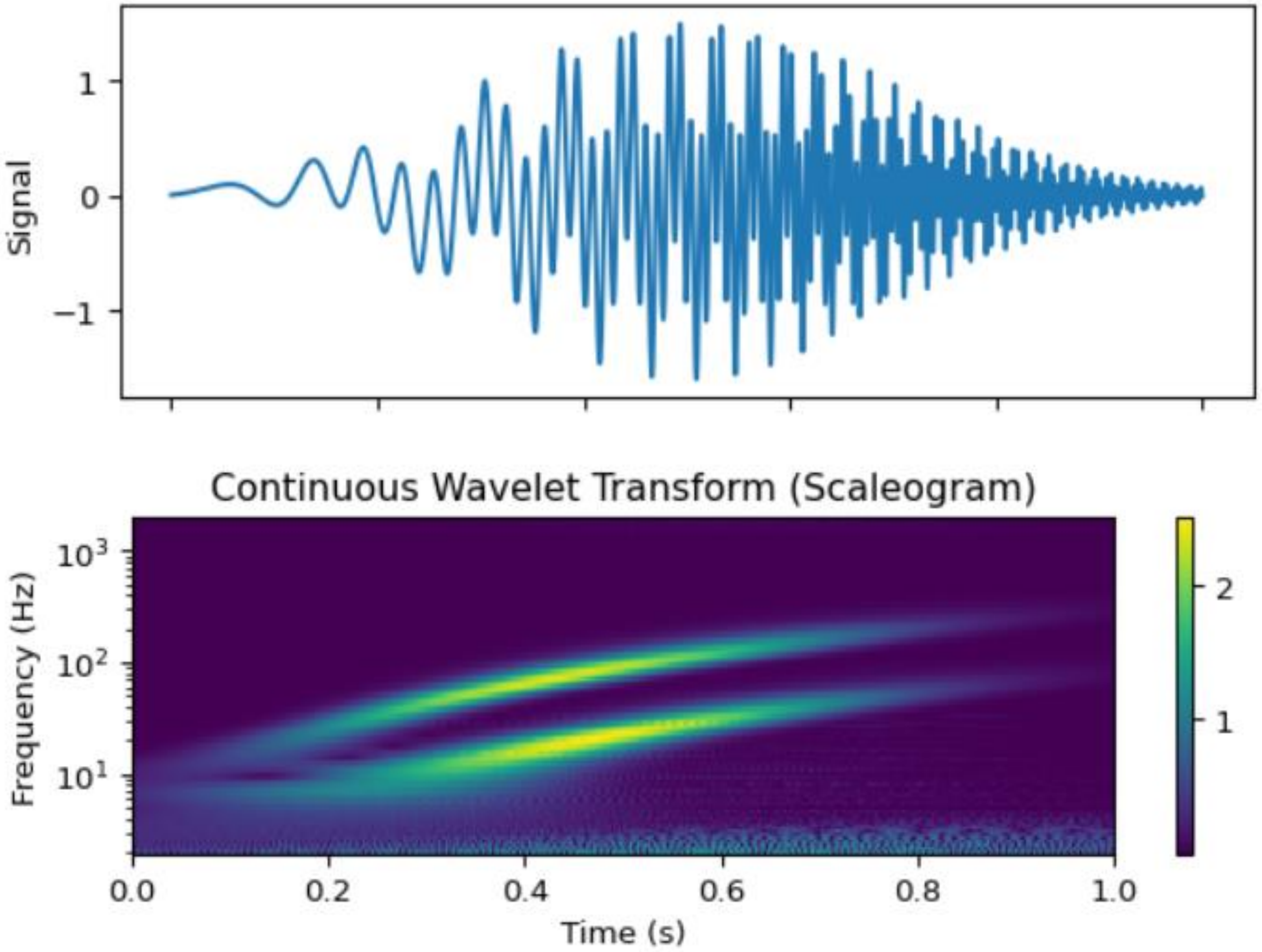
Dataset 2 is ~4 days longer than dataset 1. The frequency components appear to converge on 24-hours, and a ~12-hour component appears in the top five! Atmospheric pressure has a semi-diurnal (12-hr) variation. The 551-hr peak is likely due to aliasing.

What if 5-minute counts?



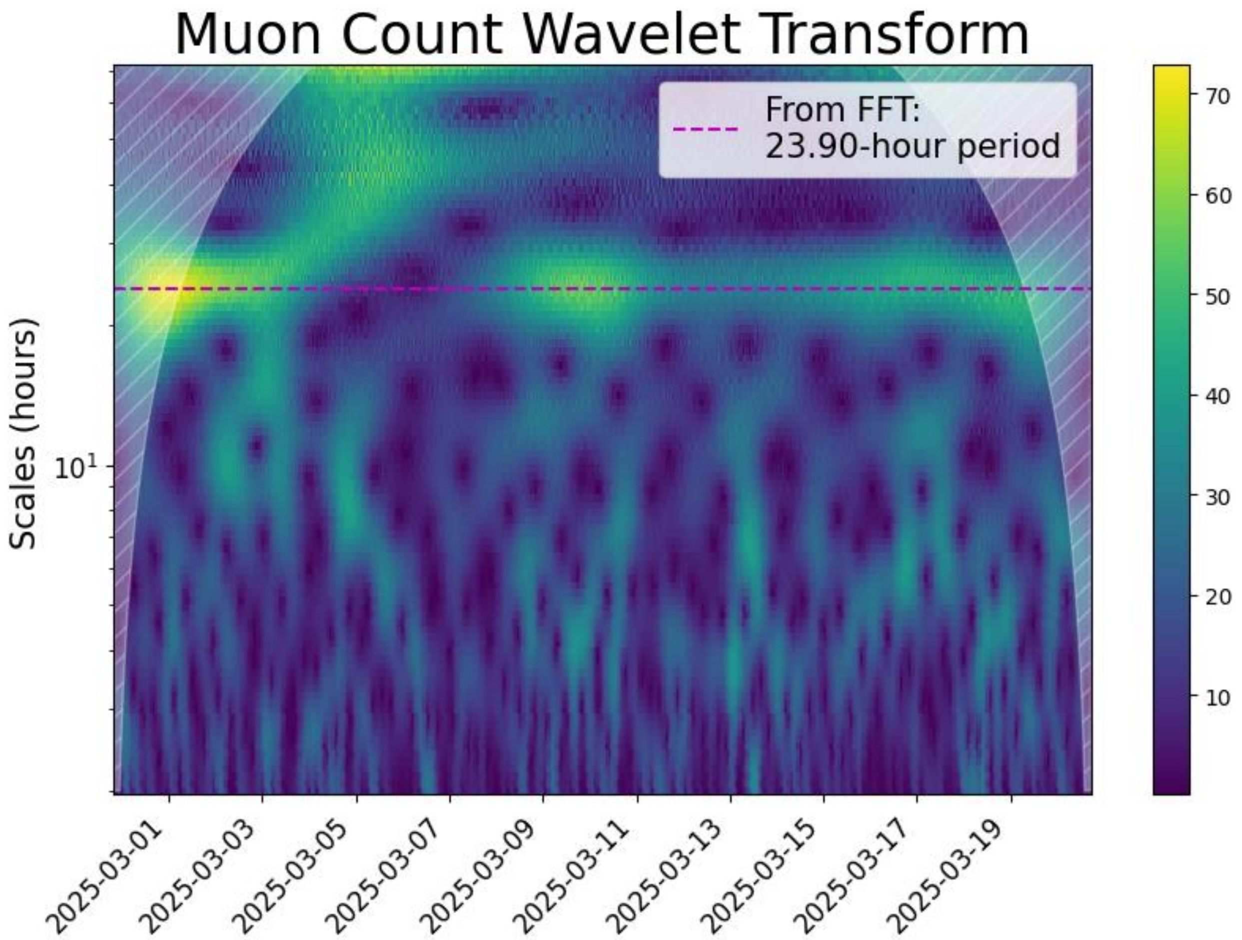
A finer sampling interval of 5-minutes improves the focus on the periodicities. Too small though, and the noise becomes statistically significant relative to the signal. Since there is still aliasing, more sophisticated removal of the background levels might be needed to refine the results.

Frequencies some-when else in time



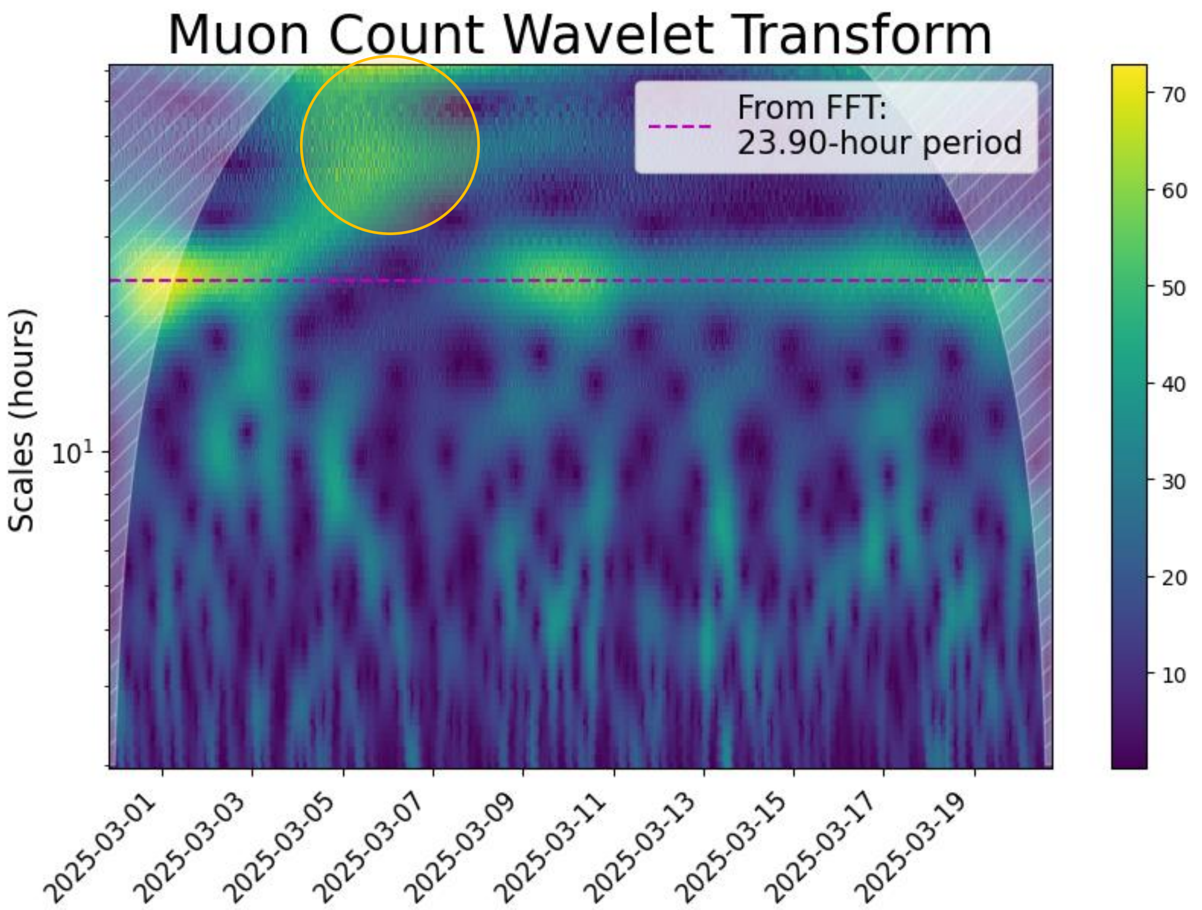
The FFT is great at finding frequencies, but assumes they are global→will not accurately reconstruct a signal if frequency components change over time. Wavelet transforms give frequencies and their *location in time.*

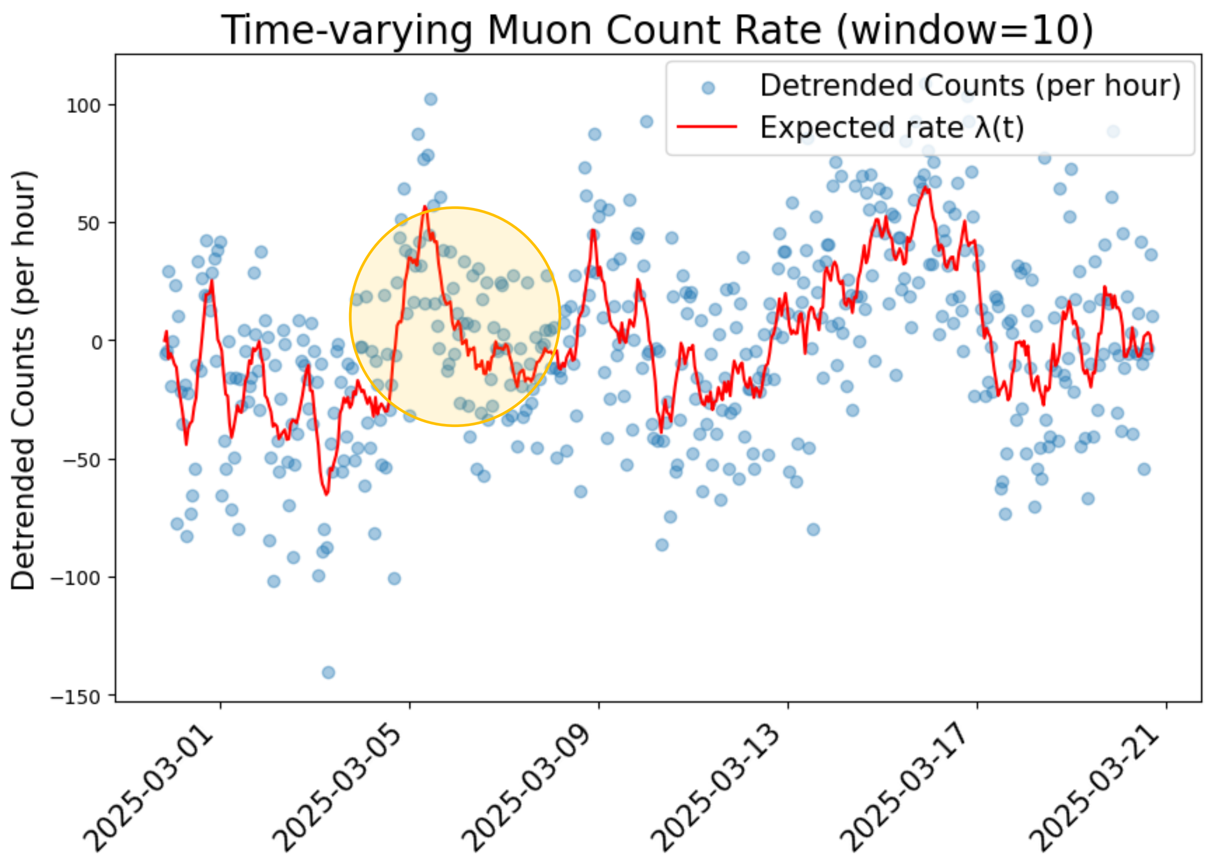
Transformuon



'Scales' determine the *width* of the analyzing wavelet. We use 100 scales ranging from 2-72 hours, log-spaced. The wavelet transform can resolve periodicities of these lengths. The cone of influence indicates where edge/boundary effects are significant. Hotspots (yellow) are analogous to peaks on the FFT.

Cross examination



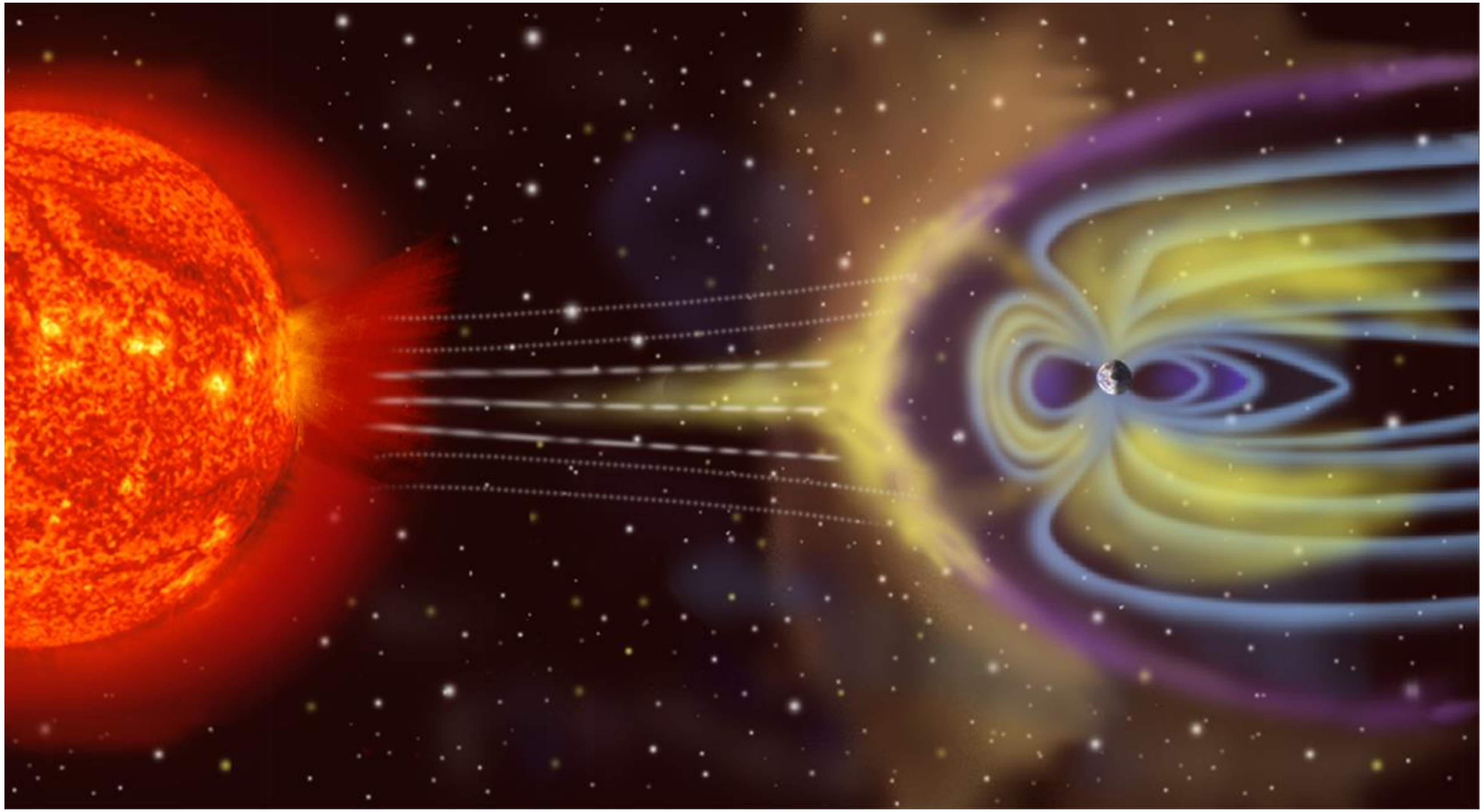


At spots of strong periodic components (left), lets look at what the muon signal was doing (right). We use a rolling average (10 hours) to estimate the count rate per hour, and denoise the signal a little.

Did something happen, spaceweather.com?

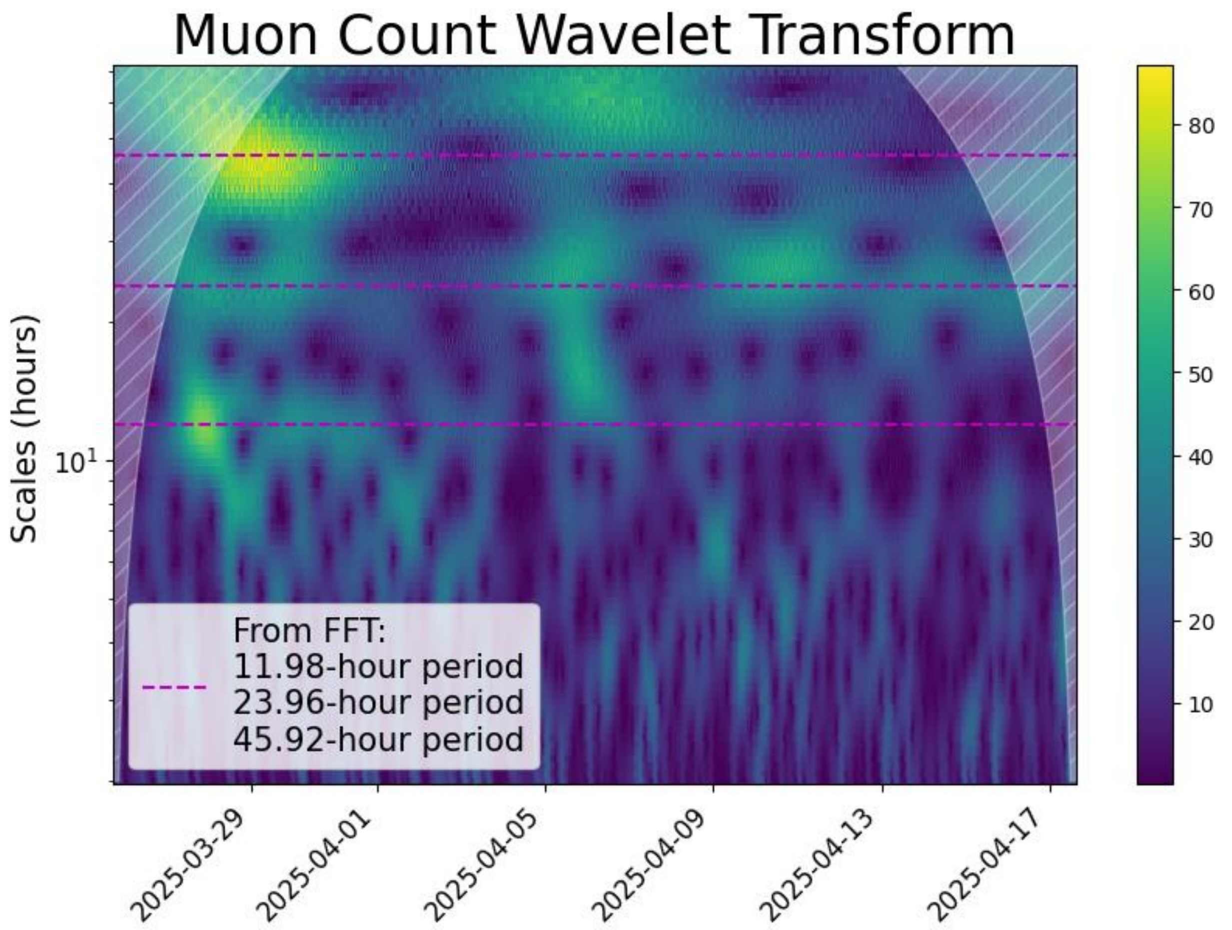
March 5, 2025

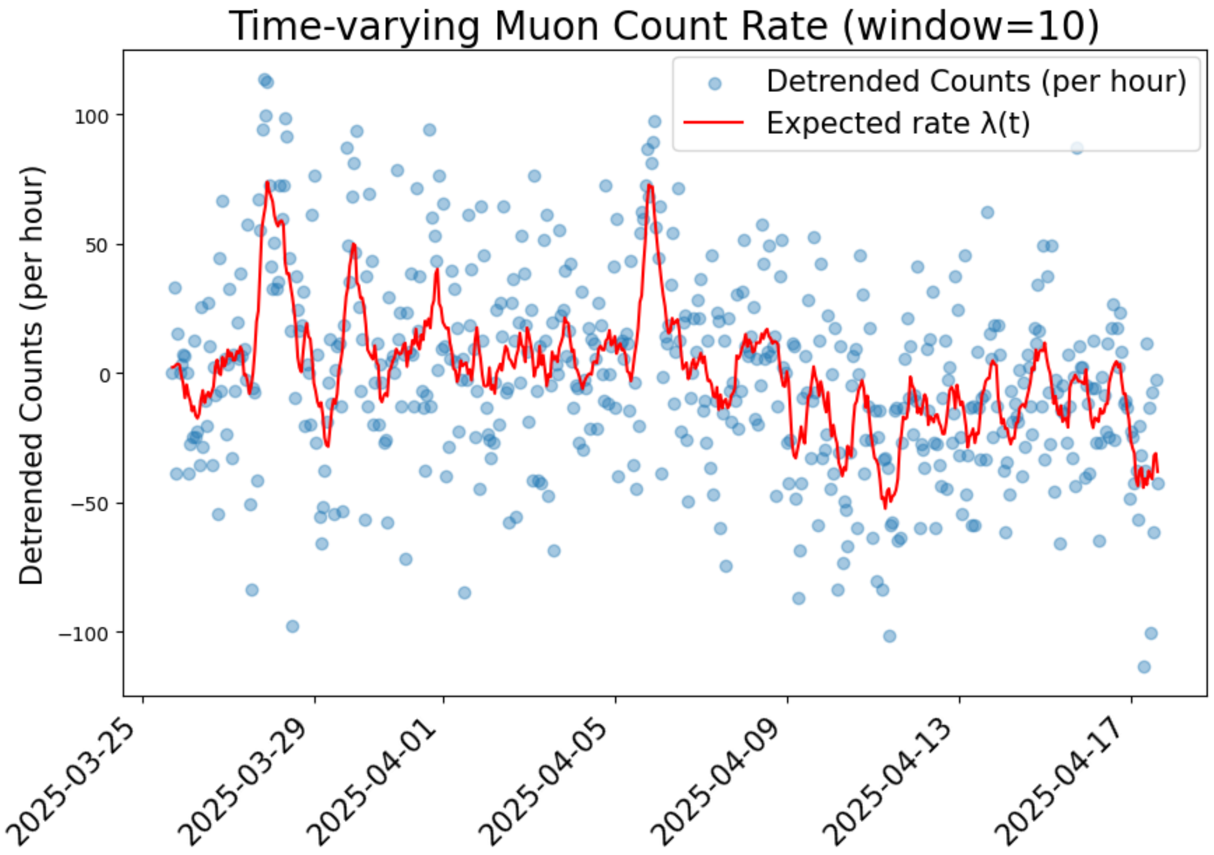
*"**MINOR GEOMAGNETIC STORM WATCH**: Minor G1-class geomagnetic storms are possible on March 5th in response to a CME, which passed near our planet during the late hours of March 4th. The CME did not deliver a direct hit. Instead, ripples of solar wind are washing over our planet like waves from a moving boat. This could be enough to cause a minor storm."*



Coronal mass ejections (CMEs) are known to cause *Forbush decreases*, which are decreases in CR flux due to increased deflection and diffusion of the CRs by the impacting CME.

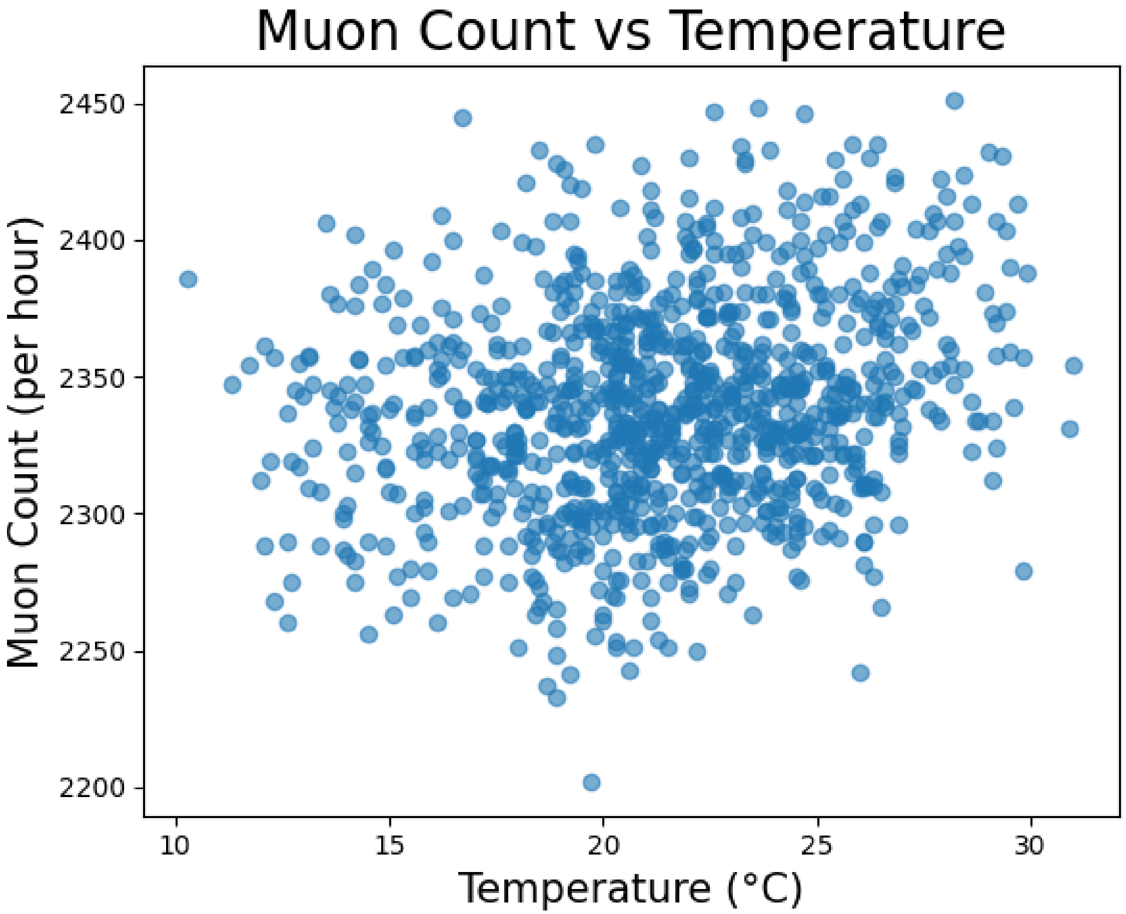
Dataset 2

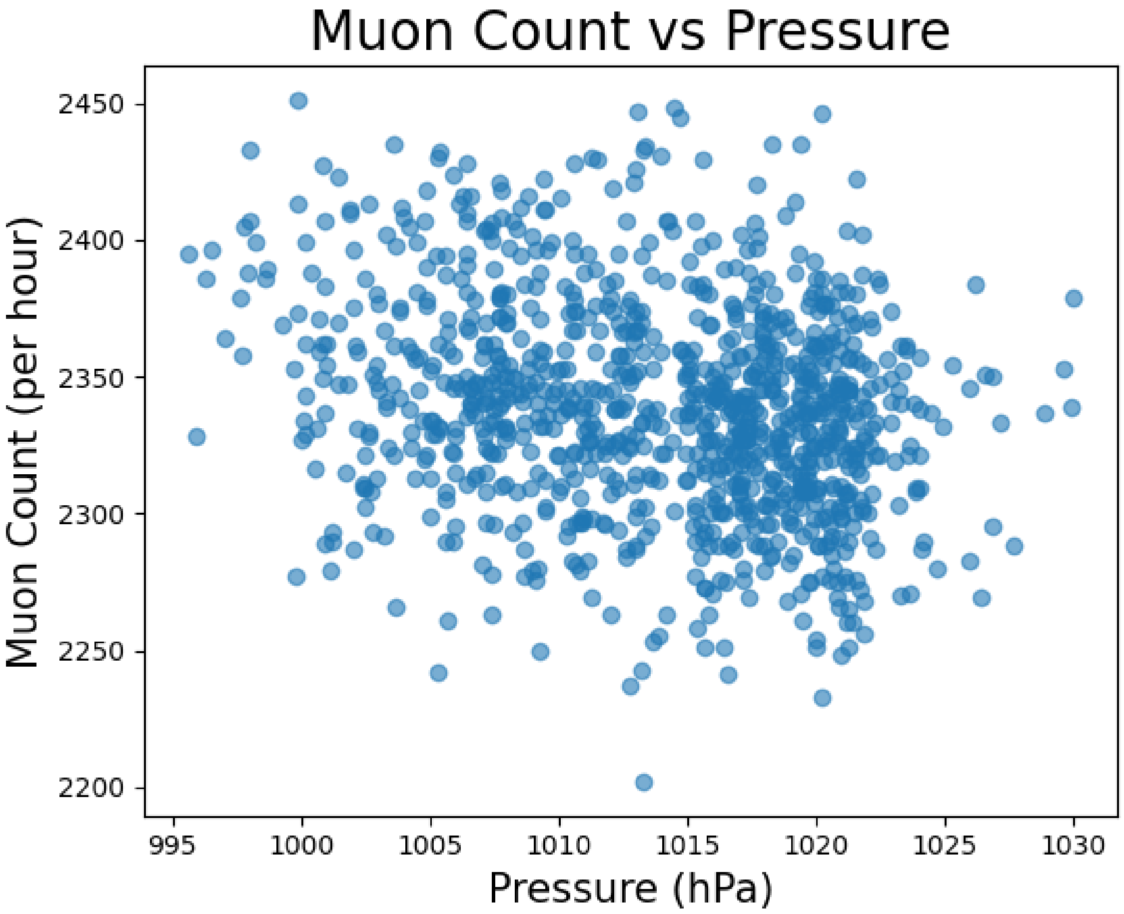




Stare at these for a bit, im gonna go take a nap, let me know what you come up with.

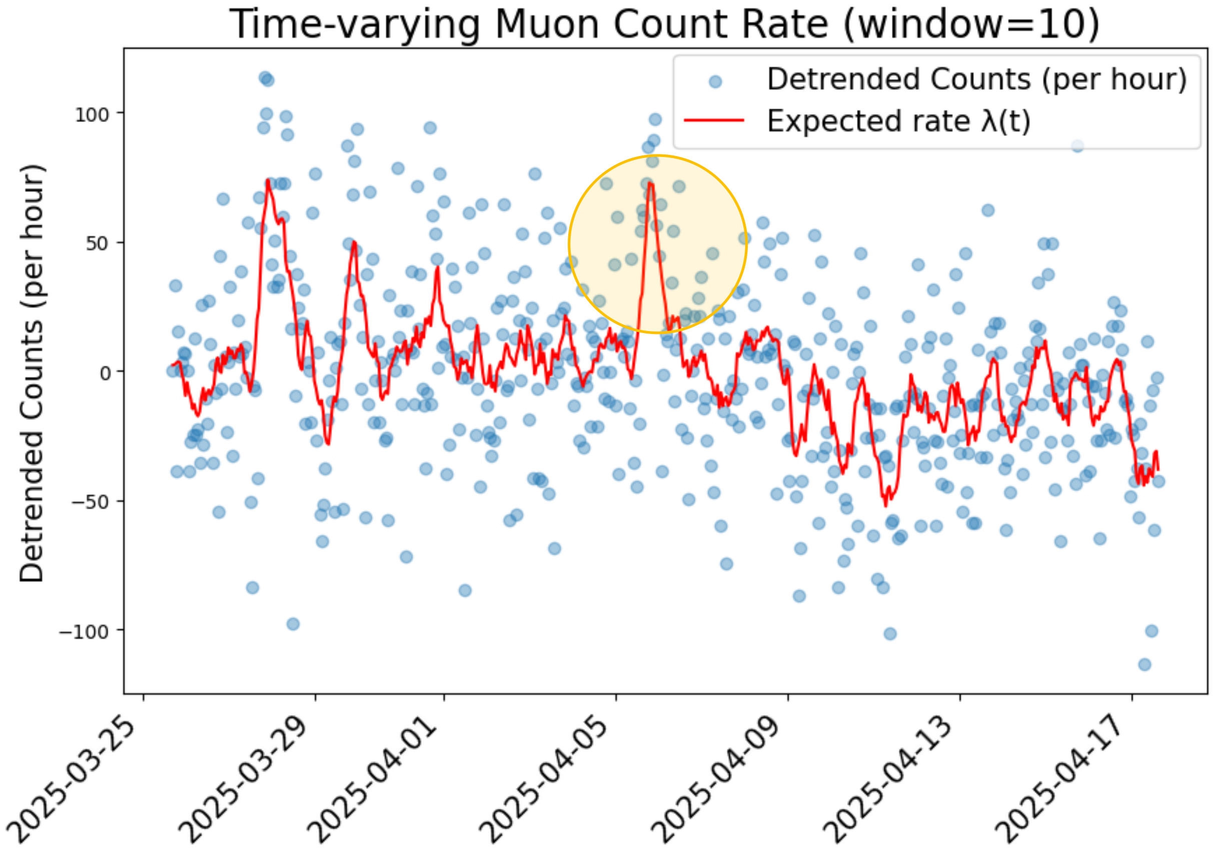
Atmospheric effects?

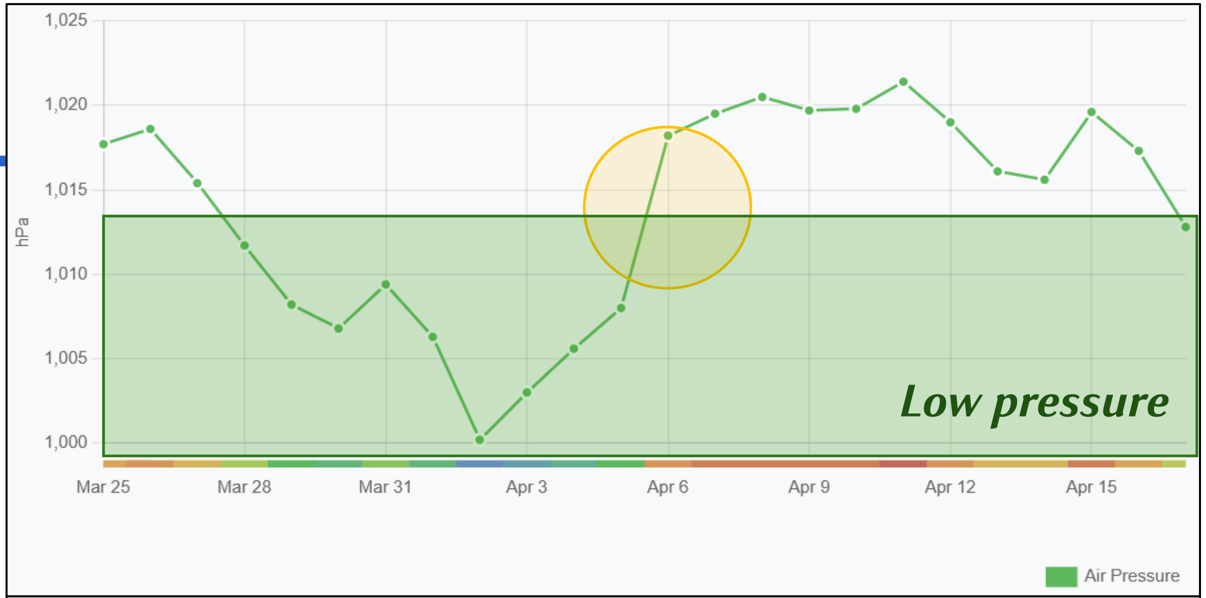


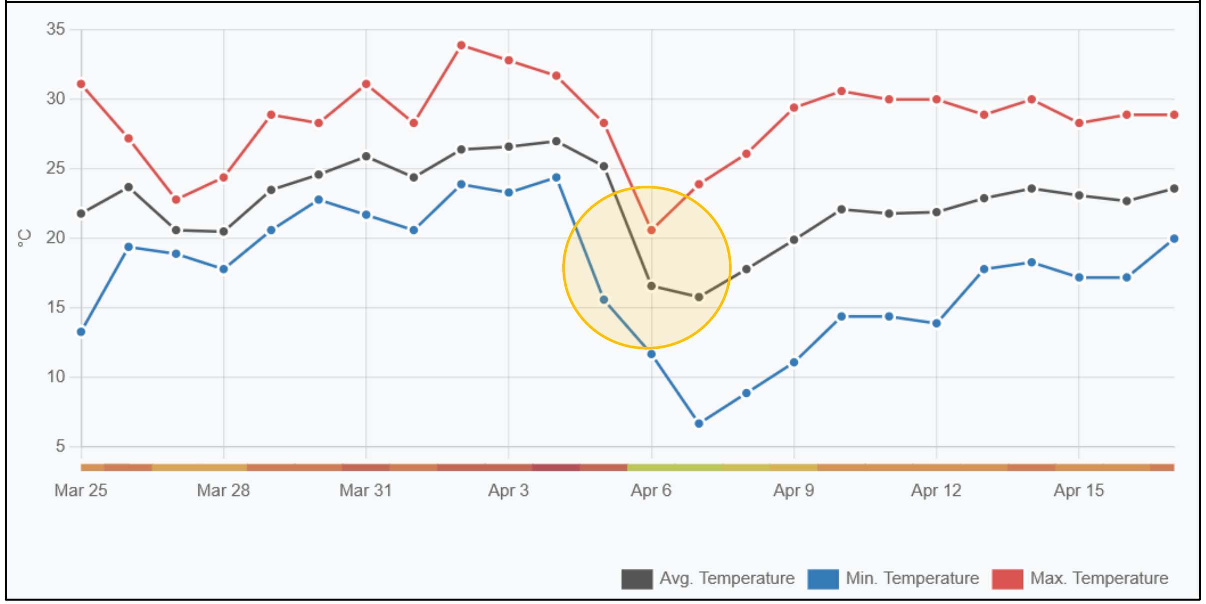


Pressure might be said to dominate in modulating our MCs. Weather data fetched from Open-Meteo.

Let’s try the weather



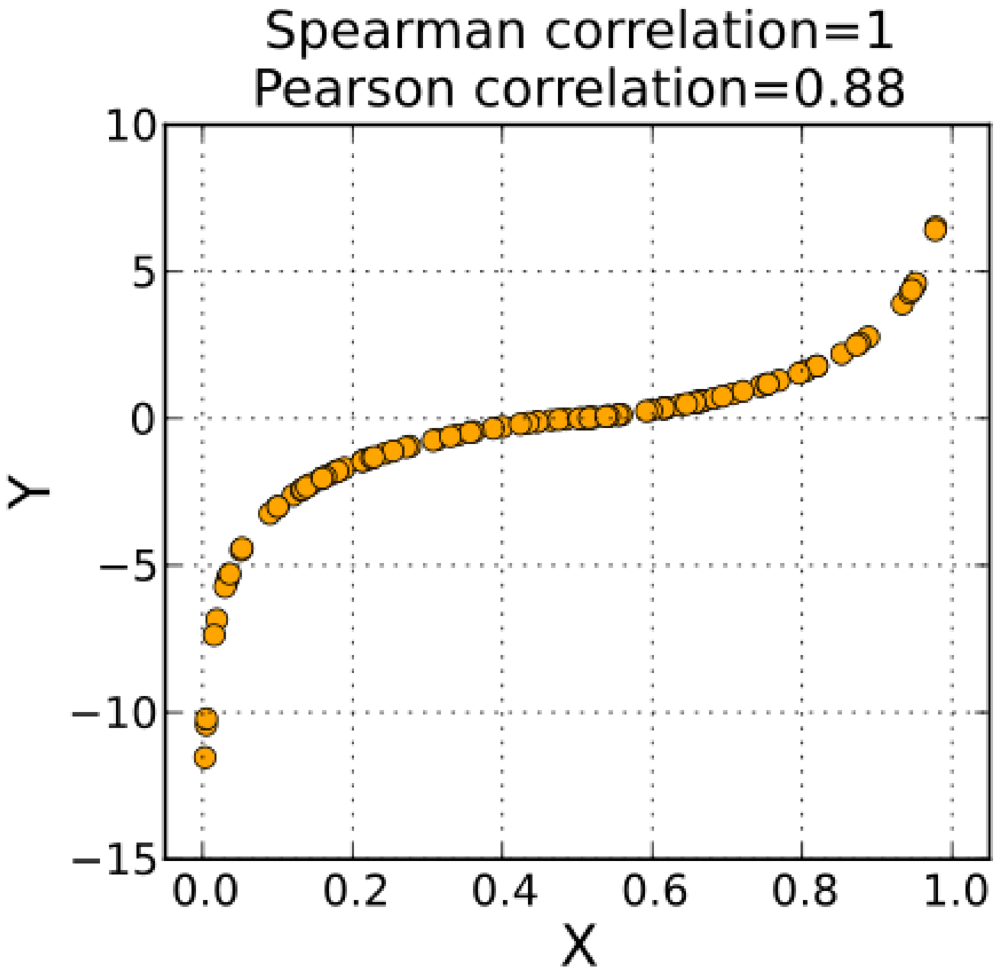




Positive muon count spikes might be from special atmospheric conditions. Weather charts from Meteostat for Corpus Christi International Airport location.

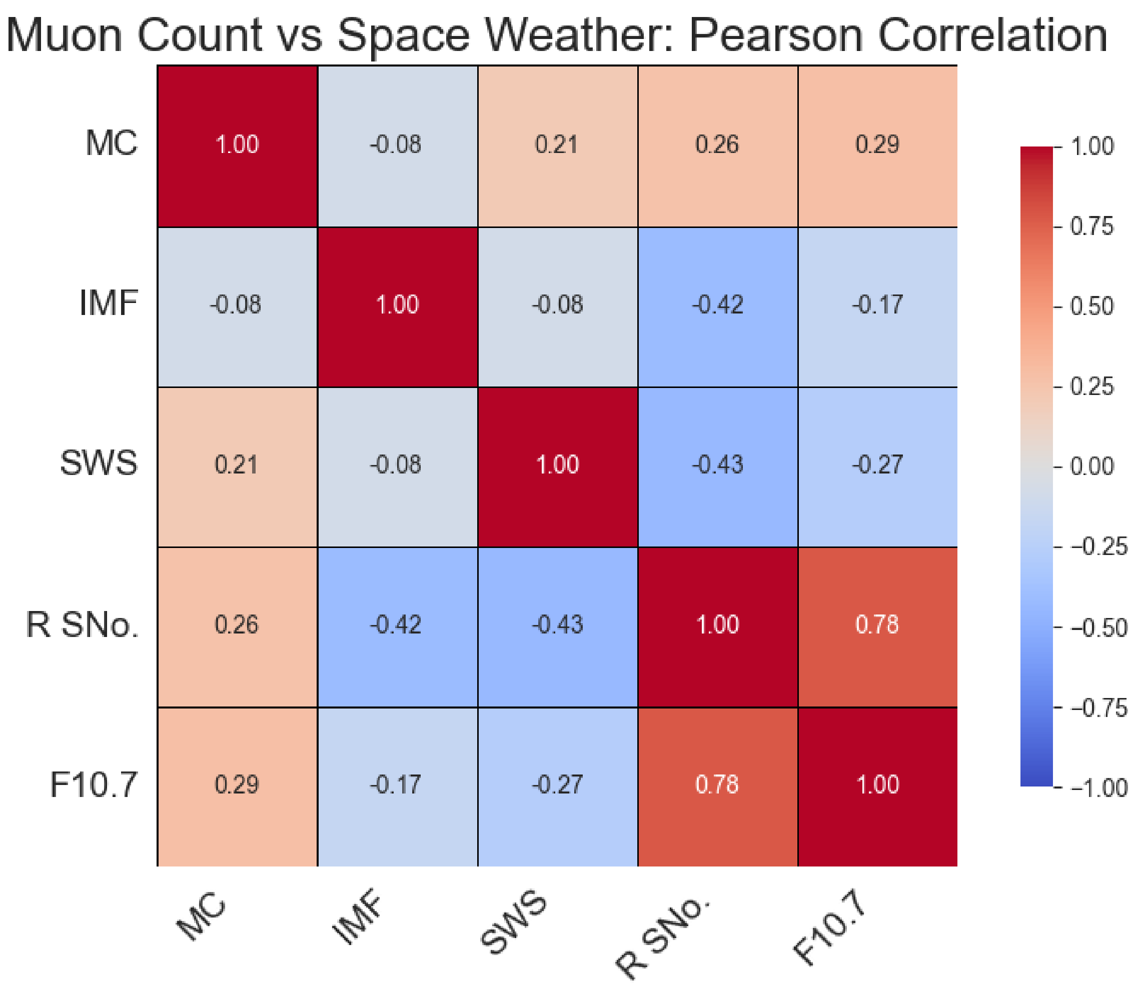
correlation analysis

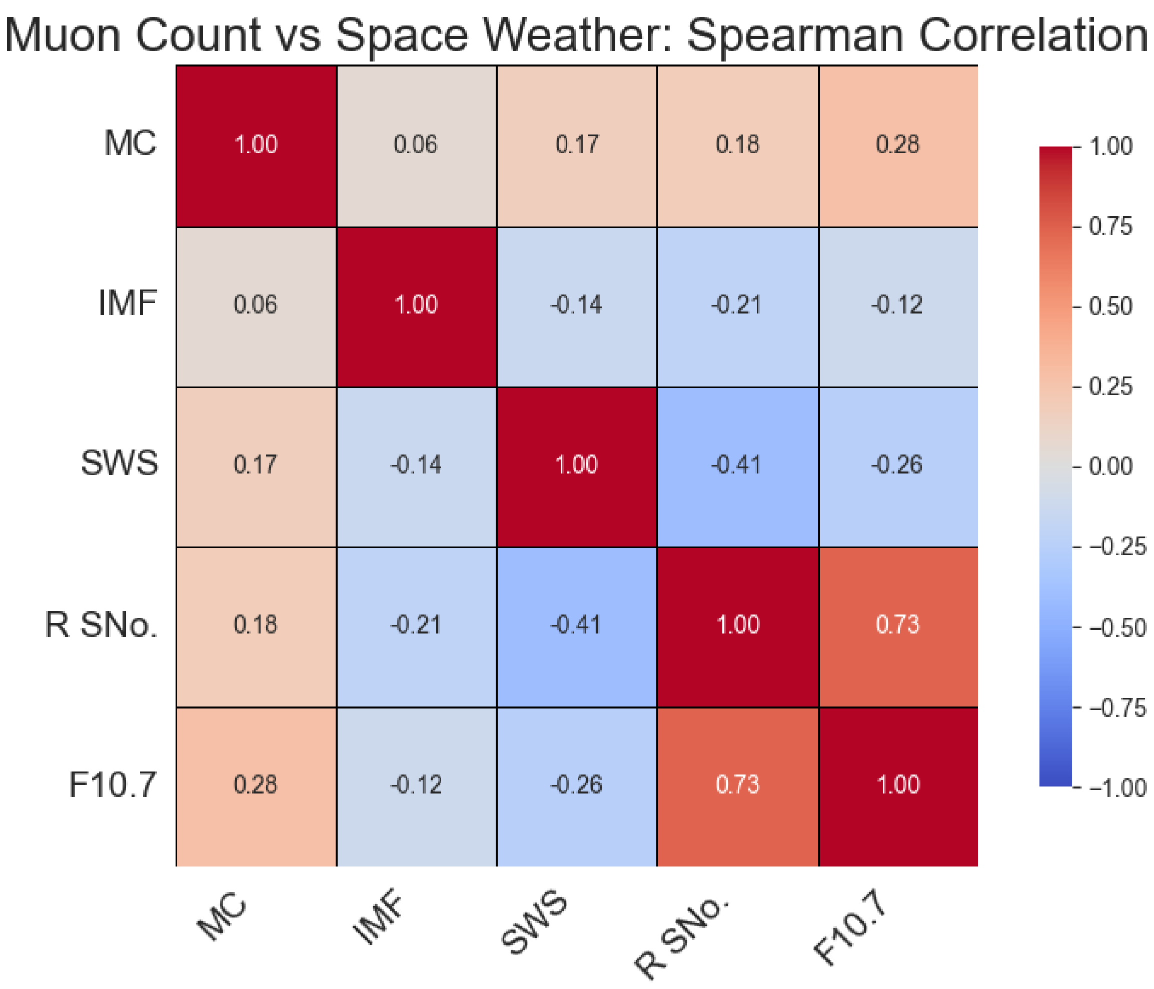
Correlation coefficients



Pearson coefficient (P) gives the strength of the linear relationship. Spearman coefficient (S) gives the strength of the linear or non-linear *monotonic* relationship.

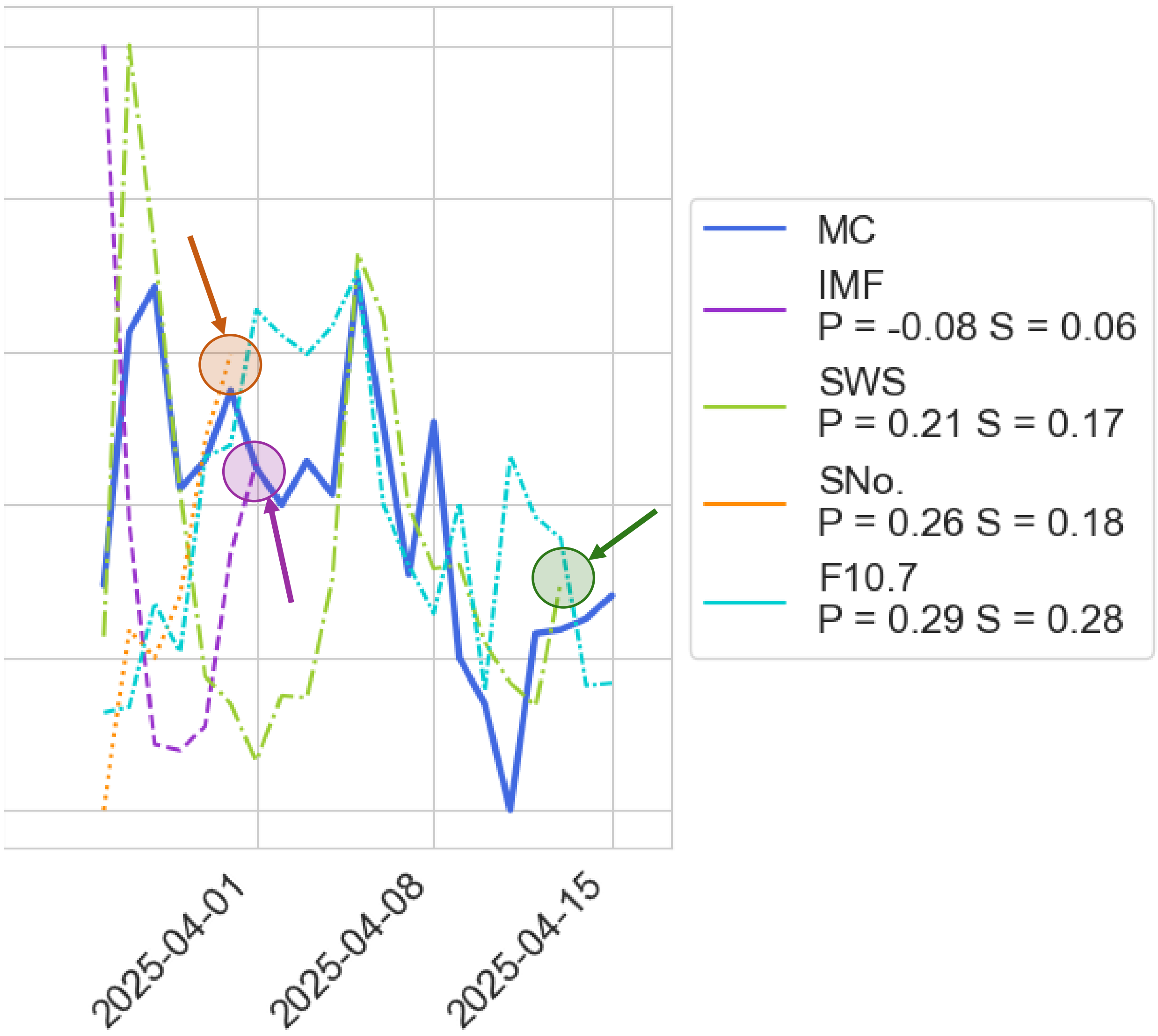
Daily data correlation heatmaps





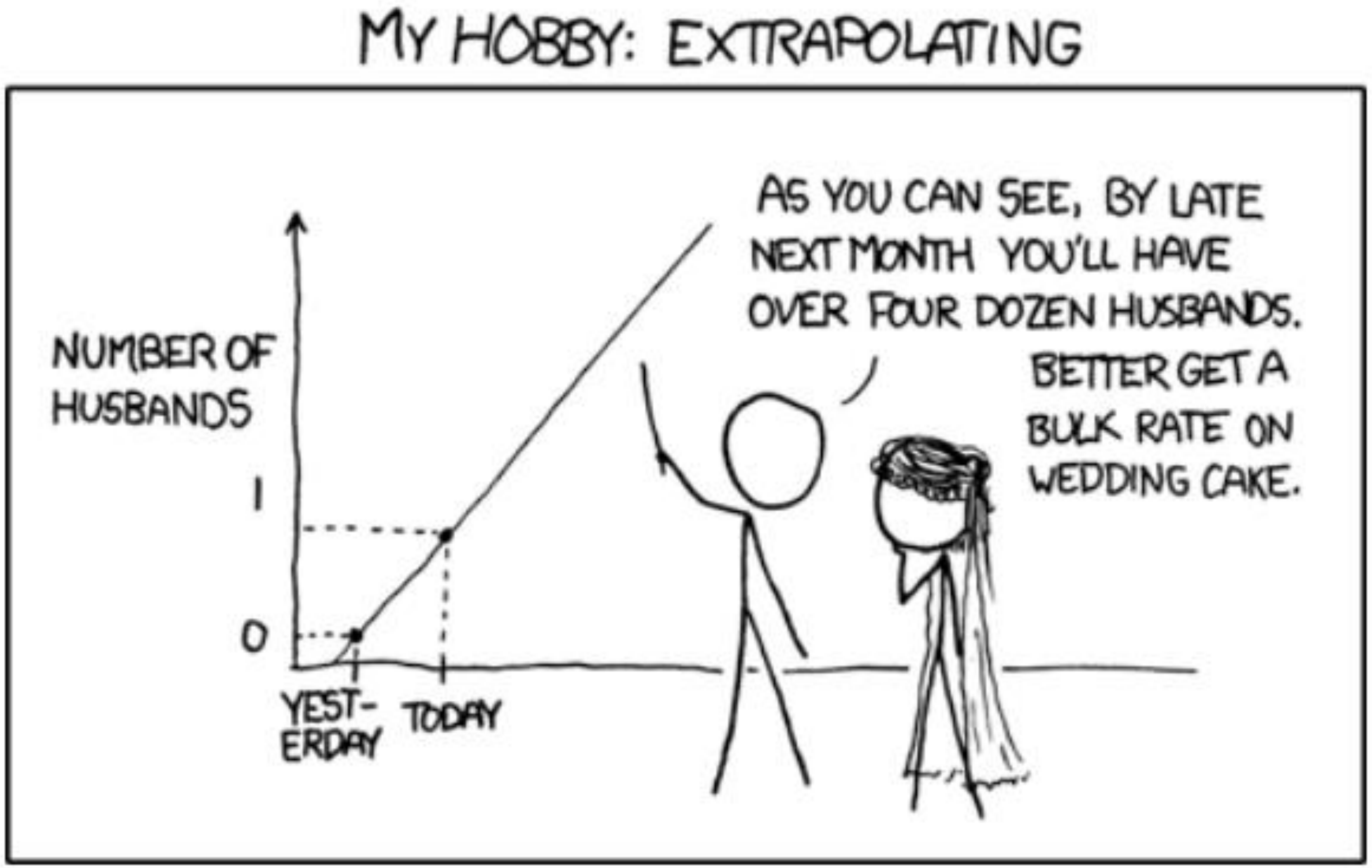
Space weather data downloaded from NASA's OMNI database. IMF: interplanetary magnetic field magnitude (nT). SWS: solar wind speed (km/s). R SNo.: R sunspot number. F10.7: solar radio flux (sfu).

Thanks, NASA



2025	83	0	9.8	395.	79	155.1
2025	84	0	9.6	372.	77	155.3
2025	85	0	16.8	471.	46	151.6
2025	86	0	8.8	750.	84	152.0
2025	87	0	5.1	654.	78	159.8
2025	88	0	5.0	539.	91	156.1
2025	89	0	5.4	452.	124	170.7
2025	90	0	8.3	439.	142	171.6
2025	91	0	9.8	412.	999	181.7
2025	92	0	999.9	443.	999	179.8
2025	93	0	999.9	442.	999	178.4
2025	94	0	999.9	498.	999	180.5
2025	95	0	999.9	652.	999	184.6
2025	96	0	999.9	622.	999	167.2
2025	97	0	999.9	532.	999	162.6
2025	98	0	999.9	503.	999	159.0
2025	99	0	999.9	505.	999	167.3
2025	100	0	999.9	468.	999	153.2
2025	101	0	999.9	449.	999	170.8
2025	102	0	999.9	438.	999	166.2
2025	103	0	999.9	497.	999	164.6
2025	104	0	999.9	9999.	999	153.6
2025	105	0	999.9	9999.	999	153.8

Two points make a line



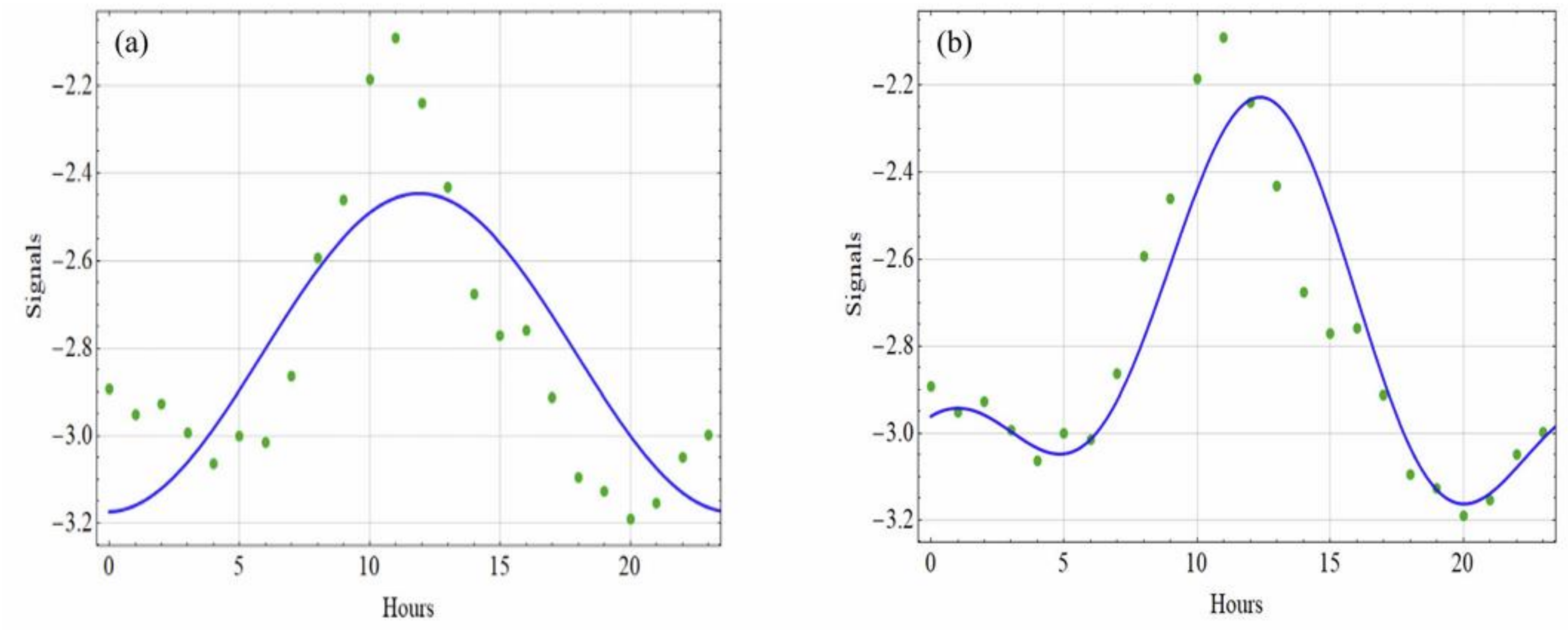
Less data makes it more likely that spurious correlations from coincidence or special circumstances not representative of physical causation will dominate calculation of the correlation strength.

Correlation stats

	Size	*P*	*P* Ci	*S*	*S* Ci
IMF	27	-0.06 ±0.18	-0.40—0.32	0.06 ±0.21	-0.34—0.47
SWS	39	0.21 ±0.13	-0.08—0.44	0.16 ±0.15	-0.15—0.45
R SNo.	26	0.26 ±0.18	-0.16—0.59	0.18 ±0.21	-0.29—0.54
F10.7	41	0.29 ±0.13	0.03—0.52	0.28 ±0.14	-0.03—0.54

There is generally higher error and lower confidence for the smaller datasets. Stronger Ps than Ss conveys the noisiness of the data. It's unlikely that the correlation with F10.7 is causal.

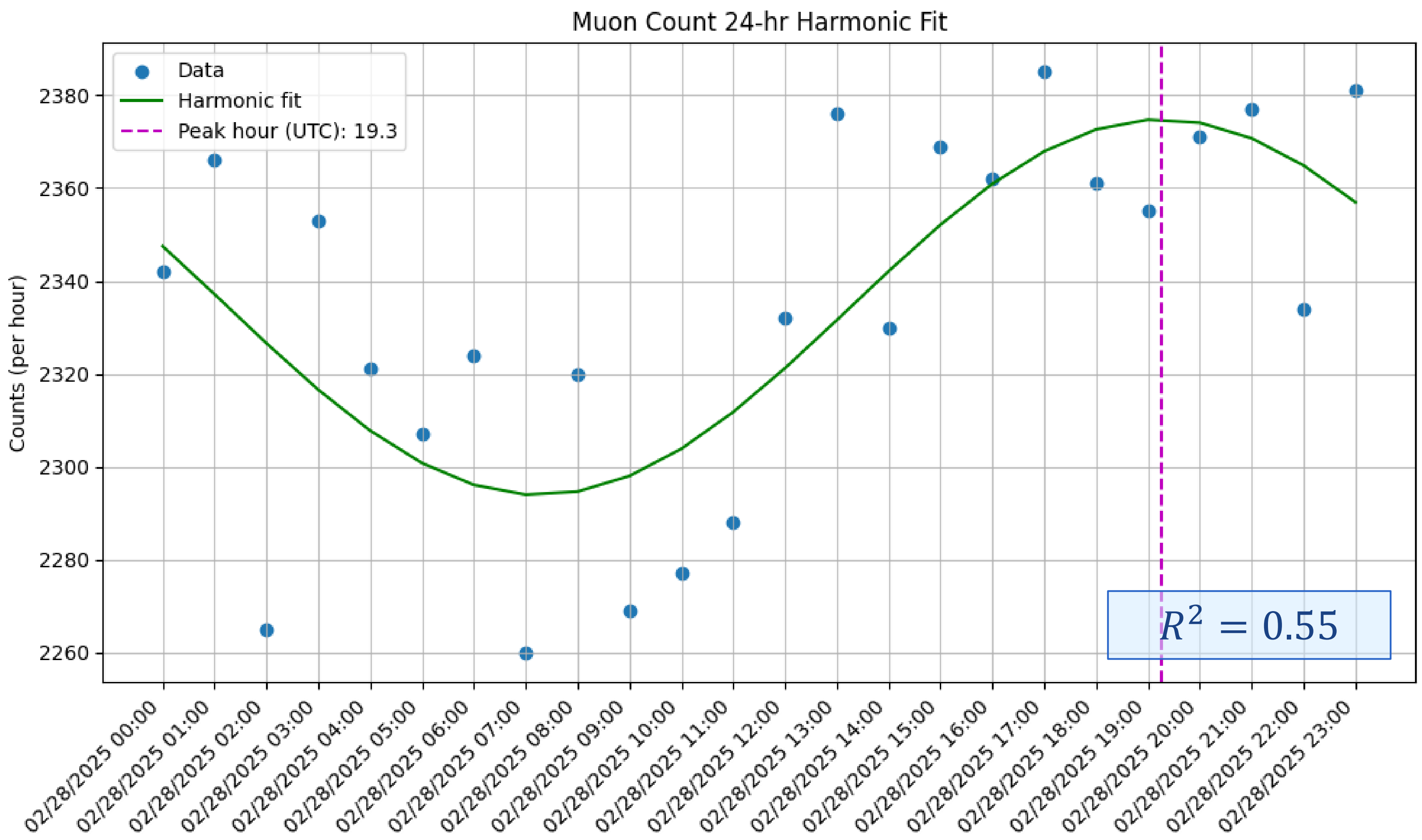
Harmonic fitting



Abdullrahman, M., A. H. Hya, and A. Aied (2025). "Exploring daily fluctuations of cosmic ray muon components at a low latitude site and their associations with space weather variables". In: *Journal of Astrophysics and Astronomy* 46.1. DOI: 10.1007/s12036-024-10034-8.

The diurnal variation of CR muons can be studied by fitting every 24 hours of data with a harmonic. Double harmonic fits are preferred but we are gonna keep it simple.

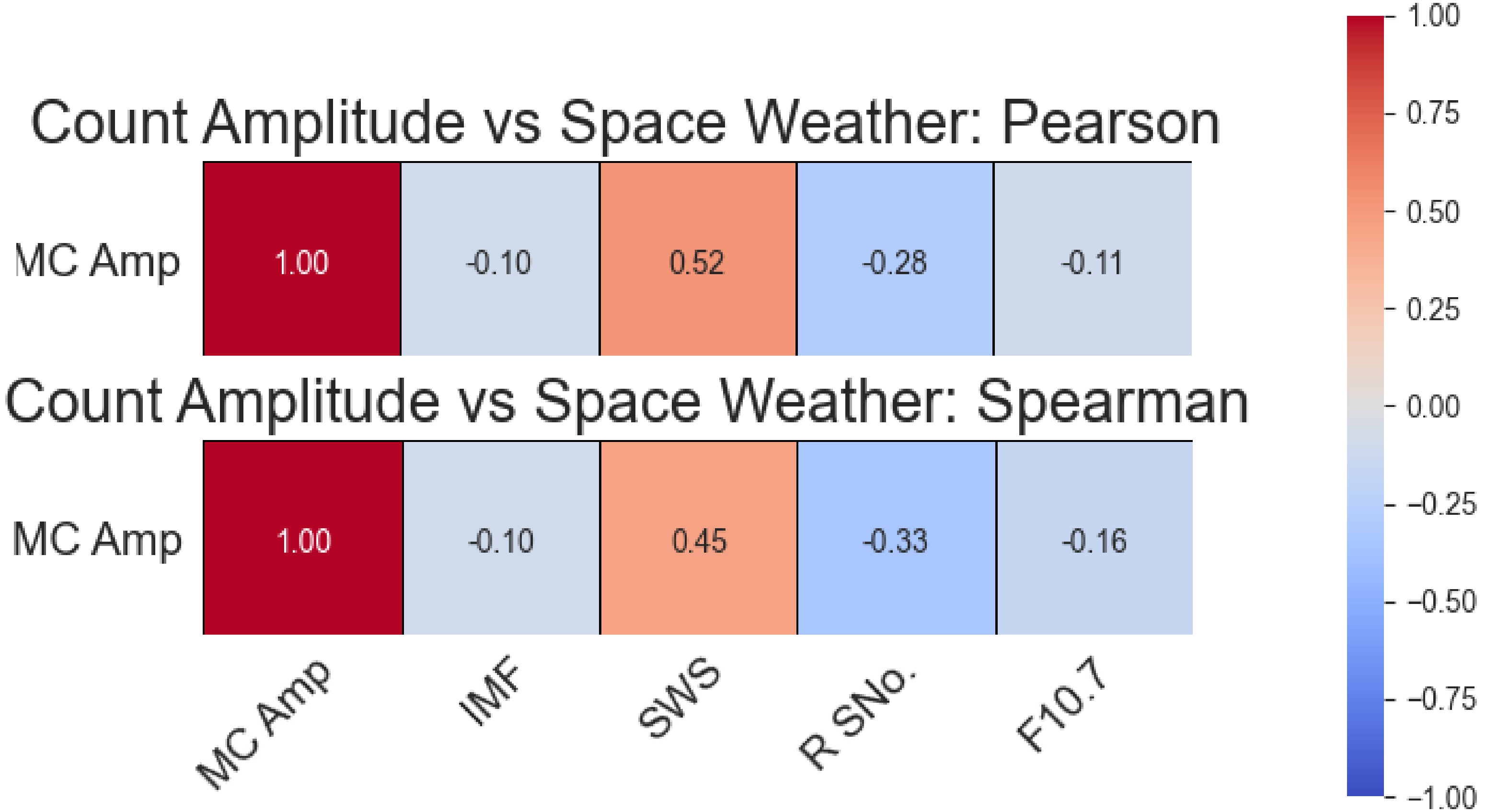
Example fit



$$f(t) = A + B\cos(\omega t + \phi)$$

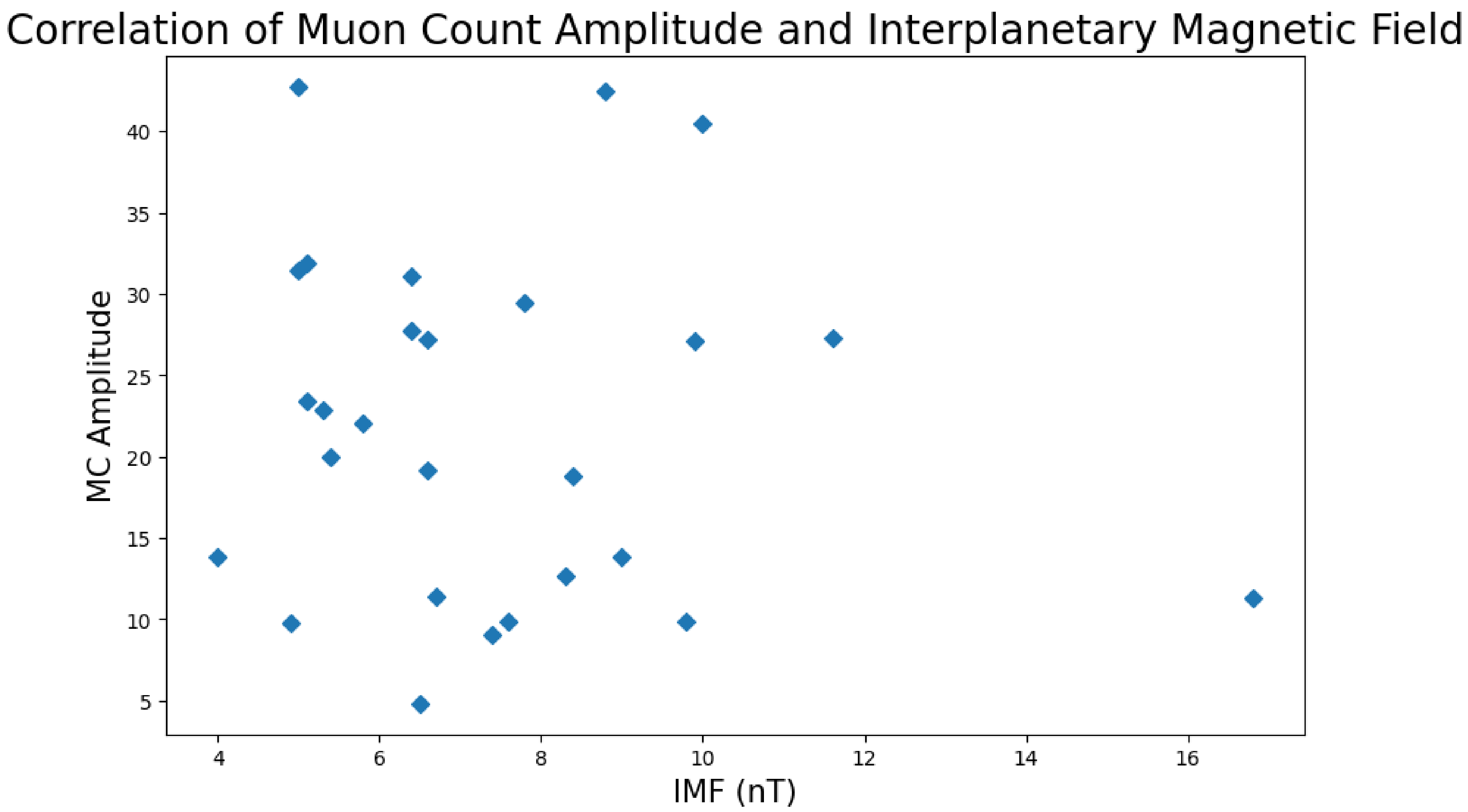
A single harmonic (sine or cosine wave) is fitted to the data so we can extract the daily amplitude B and phase ϕ→peak hour. R^2 is also calculated to measure the quality of the fit.

MC amplitude correlations



Correlations have strengthened across the board when using MC daily amplitude instead of the daily count. The amplitude acts as a measure of fluctuation throughout a day. The error is lowest, and confidence is best (not shown) for SWS, for both P and S.

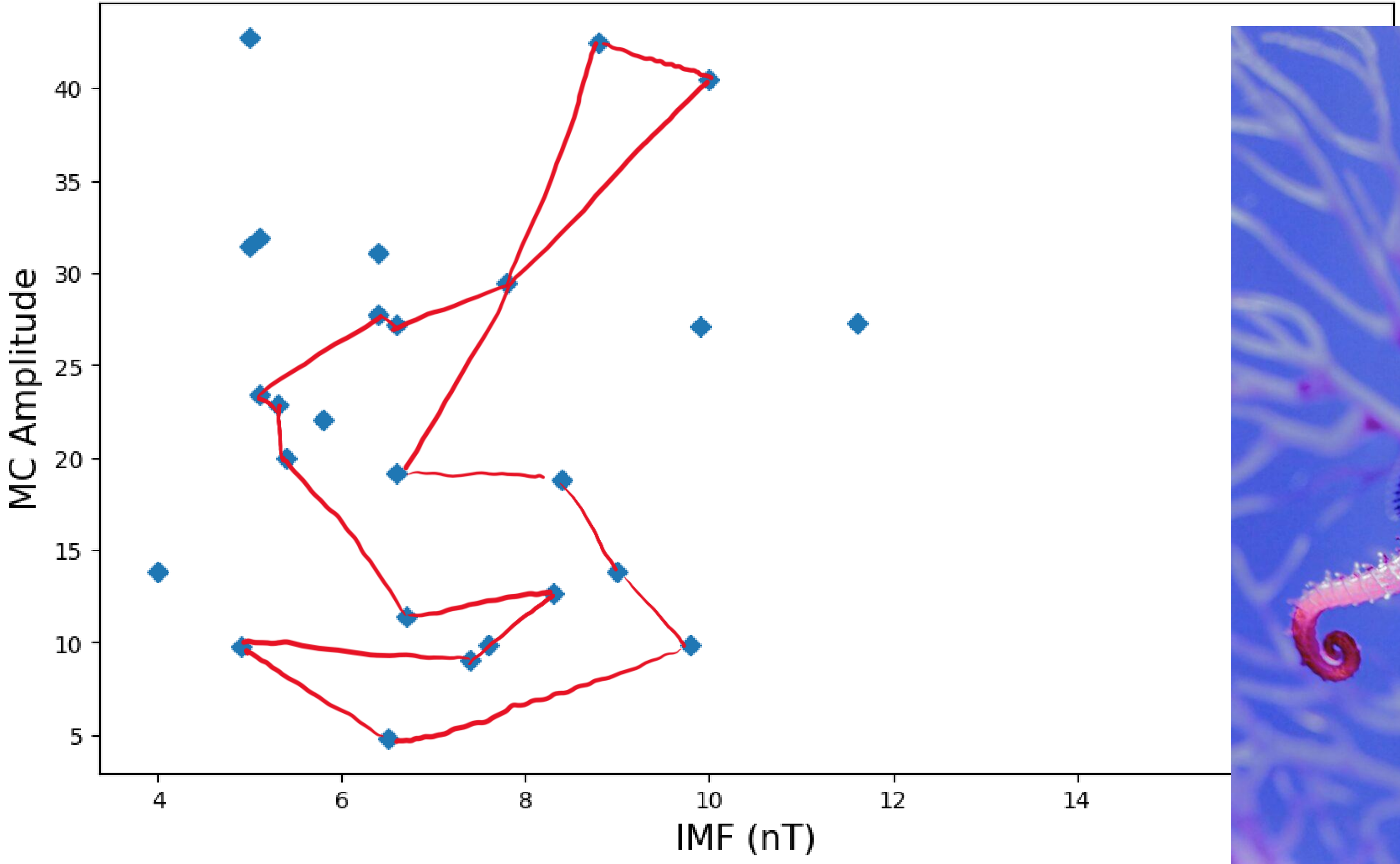
Really? No correlation with IMF?



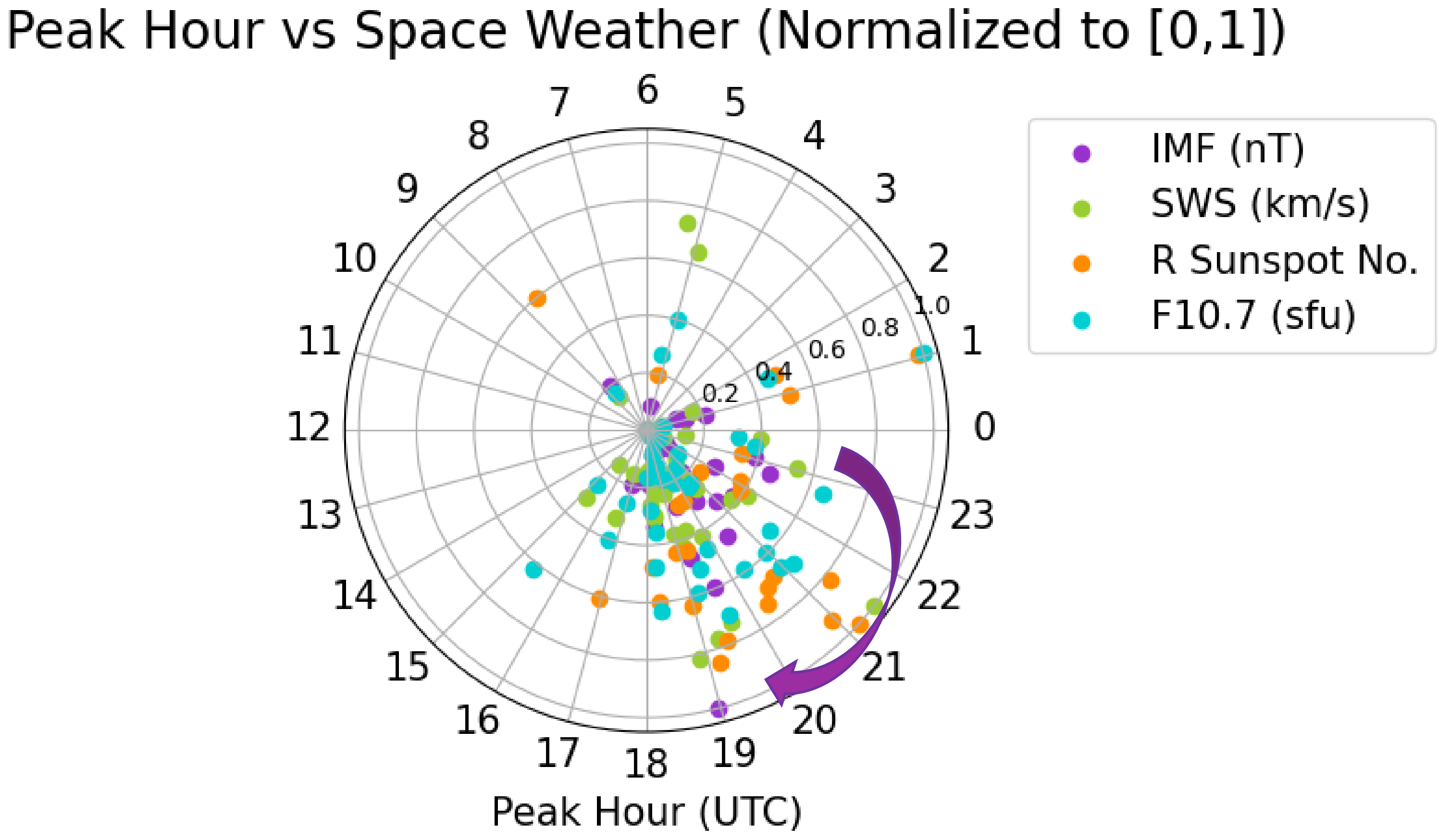
So far, I thought IMF would have a strong negative correlation, and SWS would be similar. So what's going on.

am i wrong

Correlation of Muon Count Amplitude and Interplanetary Magnetic Field



MC daily phase and space weather correlation



	r
IMF	0.30 ±0.10
SWS	0.22 ±0.12
R SNo.	0.27 ±0.13
F10.7	0.19 ±0.10

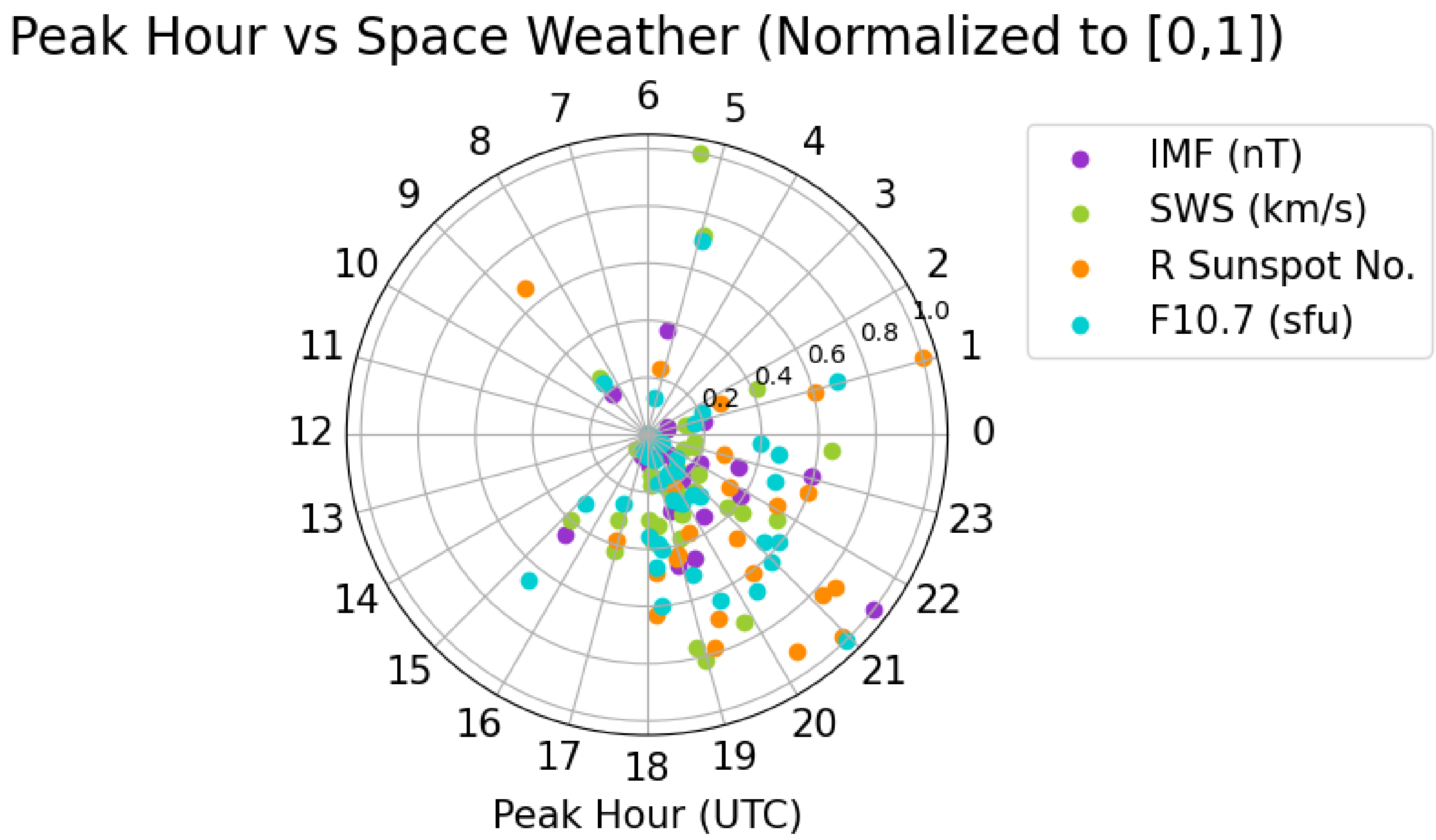
Since peak hour is a circular variable, we use circular-linear calculation of the correlation coefficient r. Maybe stronger IMF pushes the MC peak hour.

Correlating MC to space weather one day later

	$P_{current}$	$P_{shifted}$	$S_{current}$	$S_{shifted}$
IMF	-0.06 ±0.18	0.24 ±0.16	0.06 ±0.21	0.29 ±0.20
SWS	0.21 ±0.13	0.00 ±0.15	0.16 ±0.15	-0.02 ±0.15
R SNo.	0.26 ±0.18	0.19 ±0.18	0.18 ±0.21	0.12 ±0.21
F10.7	0.29 ±0.13	0.27 ±0.14	0.28 ±0.14	0.23 ±0.15

The literature says the strongest correlation with IMF is with CR flux one day later. (There is some delay in the response of CRs to changes in the heliosphere.) IMF corr is positive now? SWS corr got worse. Results were similar for the MC amplitudes. One positive result is that corrs with SNo. and F10.7 didn't change much.

Correlating MC phase one day later



	r_{avg}
IMF	0.24 ±0.13
SWS	0.36 ±0.17
R SNo.	0.25 ±0.13
F10.7	0.20 ±0.10

Correlations are pretty much the same as when unshifted. IMF correlation has weakened slightly, and SWS has strengthened slightly.

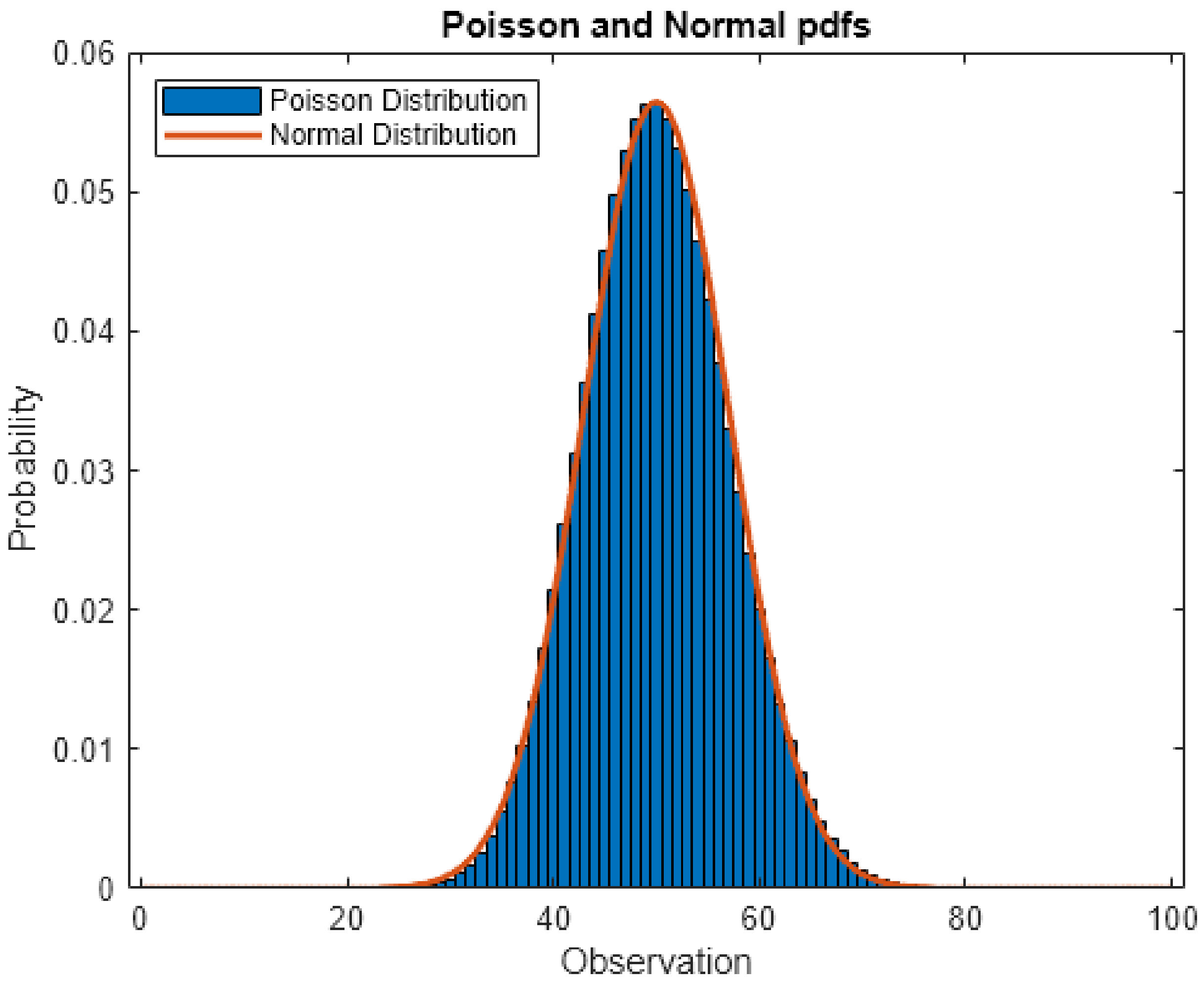
Ideas for extensions

- Correlate SWVs with atmospheric temperature and pressure
  - Verifies that SWV correlation with MC data is actually correlation with atmos data.
- Examine *hourly* correlation with space weather data.
  - Check behavior of variables during known CMEs, etc.
  - Establish the time lag of when heliospheric changes manifest in MCs.
- Extend variables to include geomagnetic activity.
  - These can help with identifying geomagnetic storm events.
  - Kp and Dst Index.
- Larger dataset, with more sophisticated background radiation removal.

thank you!

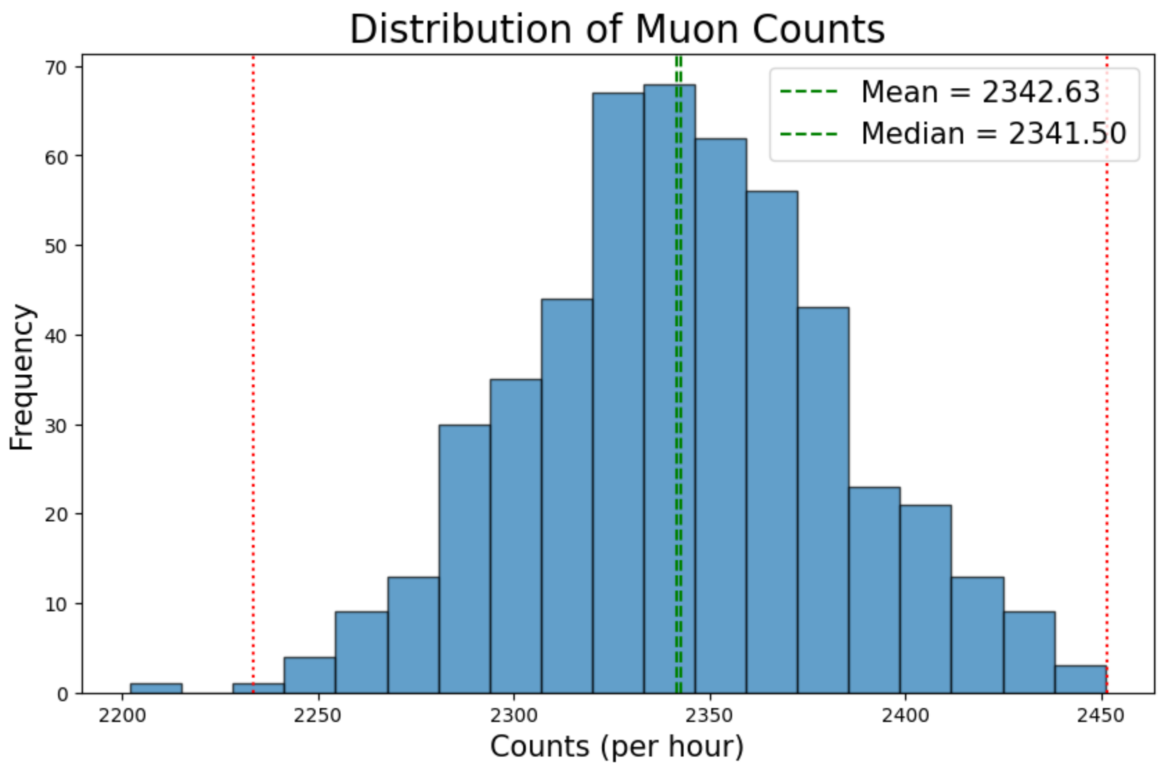
backup

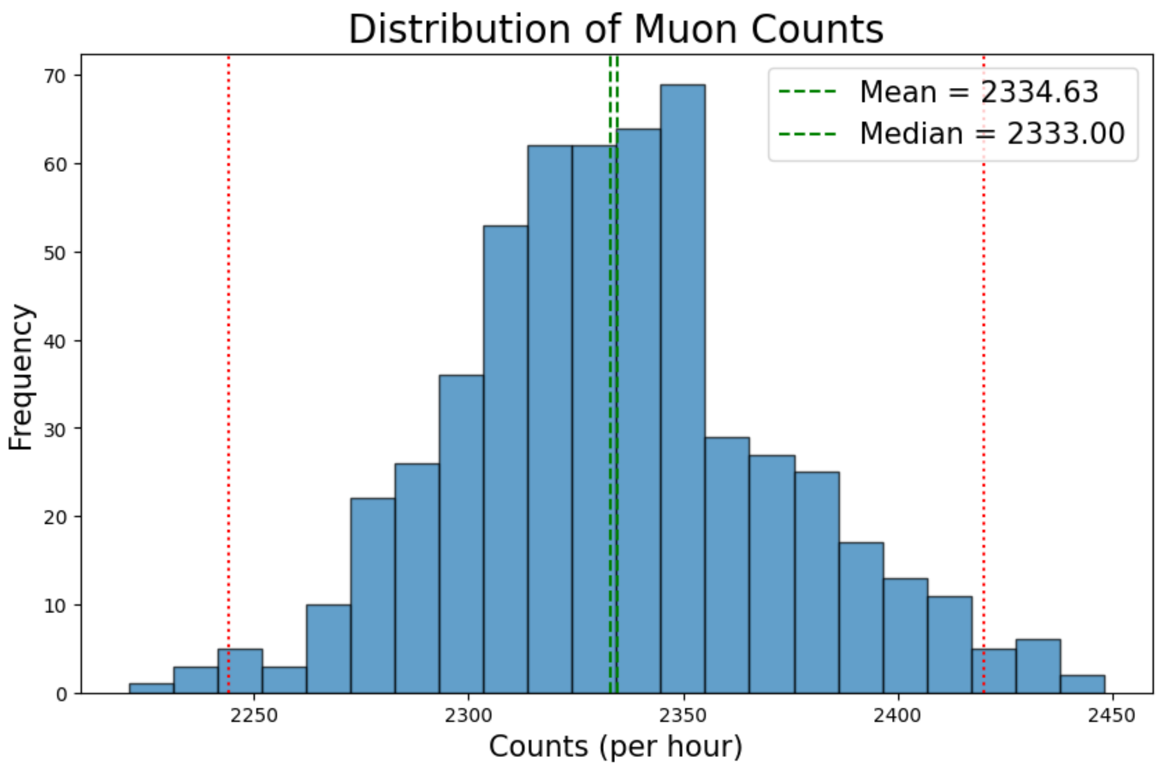
Poisson distribution



A Poisson distribution gives the probability that an event (that occurs independently with a constat average rate) will happen in a fixed interval. Similar to the normal distribution except it is discrete. E.g. typos.

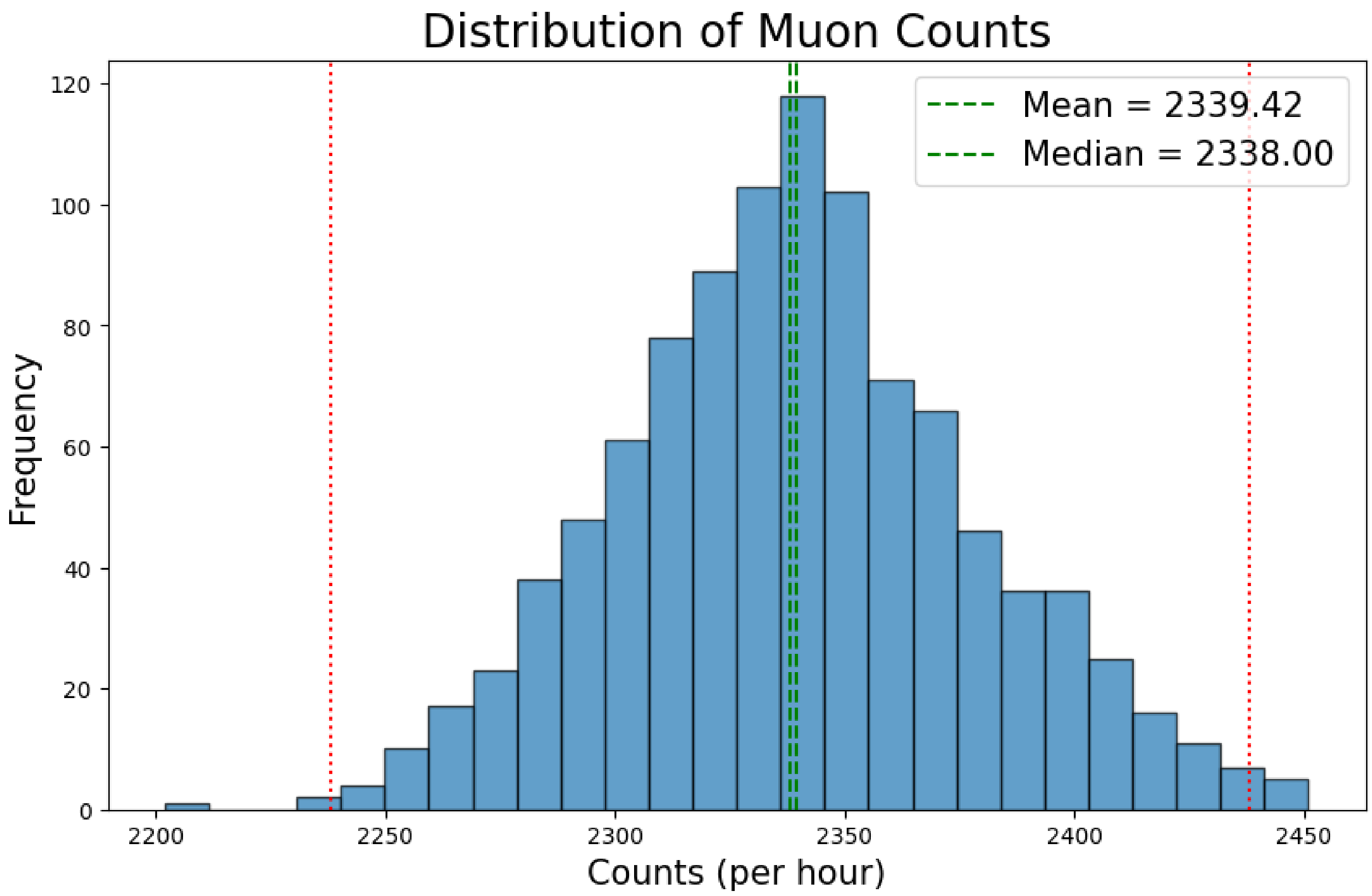
Histogram per dataset





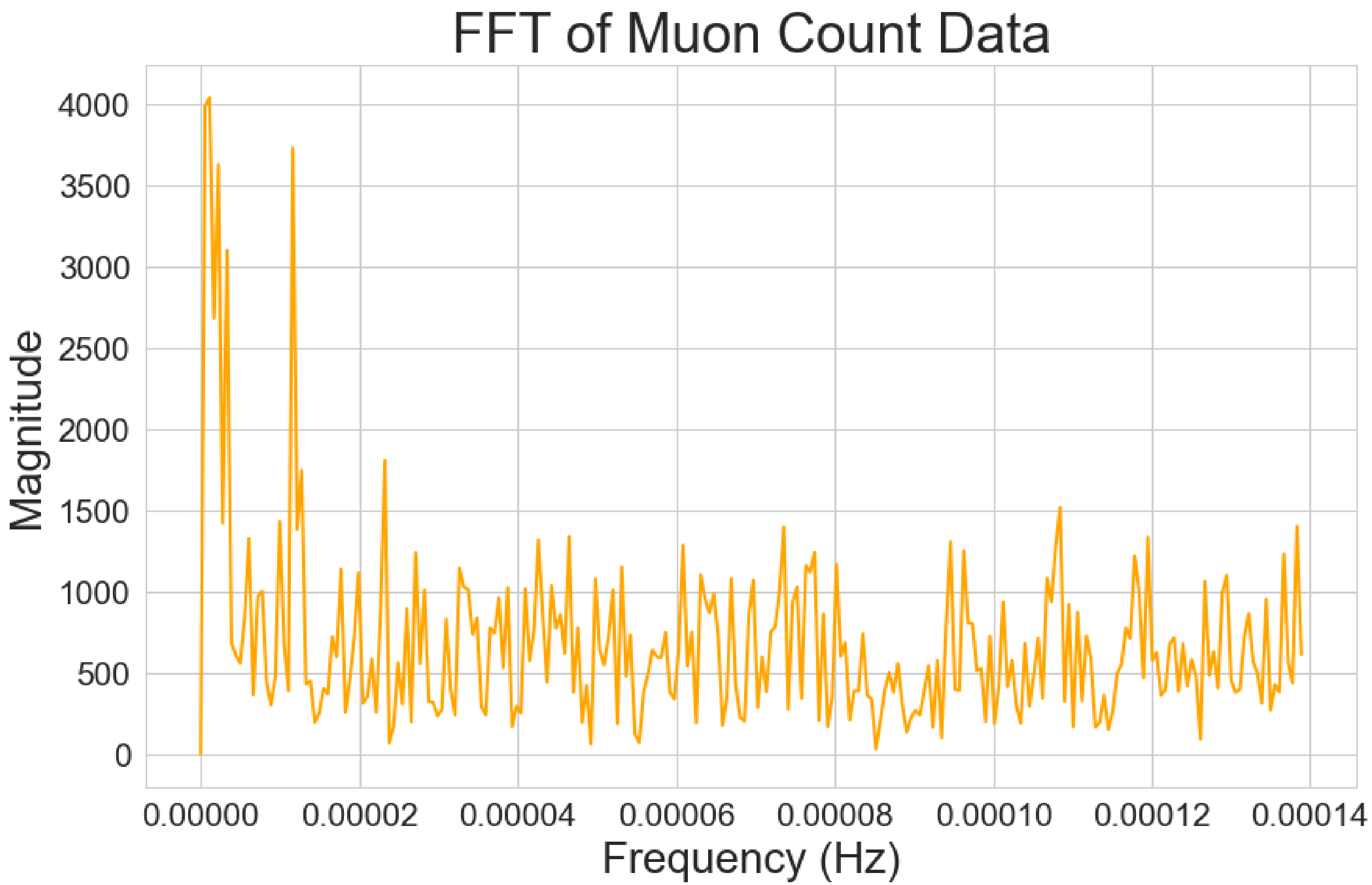
Left is 02/27 – 03/20. Right is 03/25 – 04/19. Red lines indicate outliers determined by 1.5*IQR.
The mean and median are to indicate any skew, but we see the data approximates a Poisson distribution.

Poisson muons?



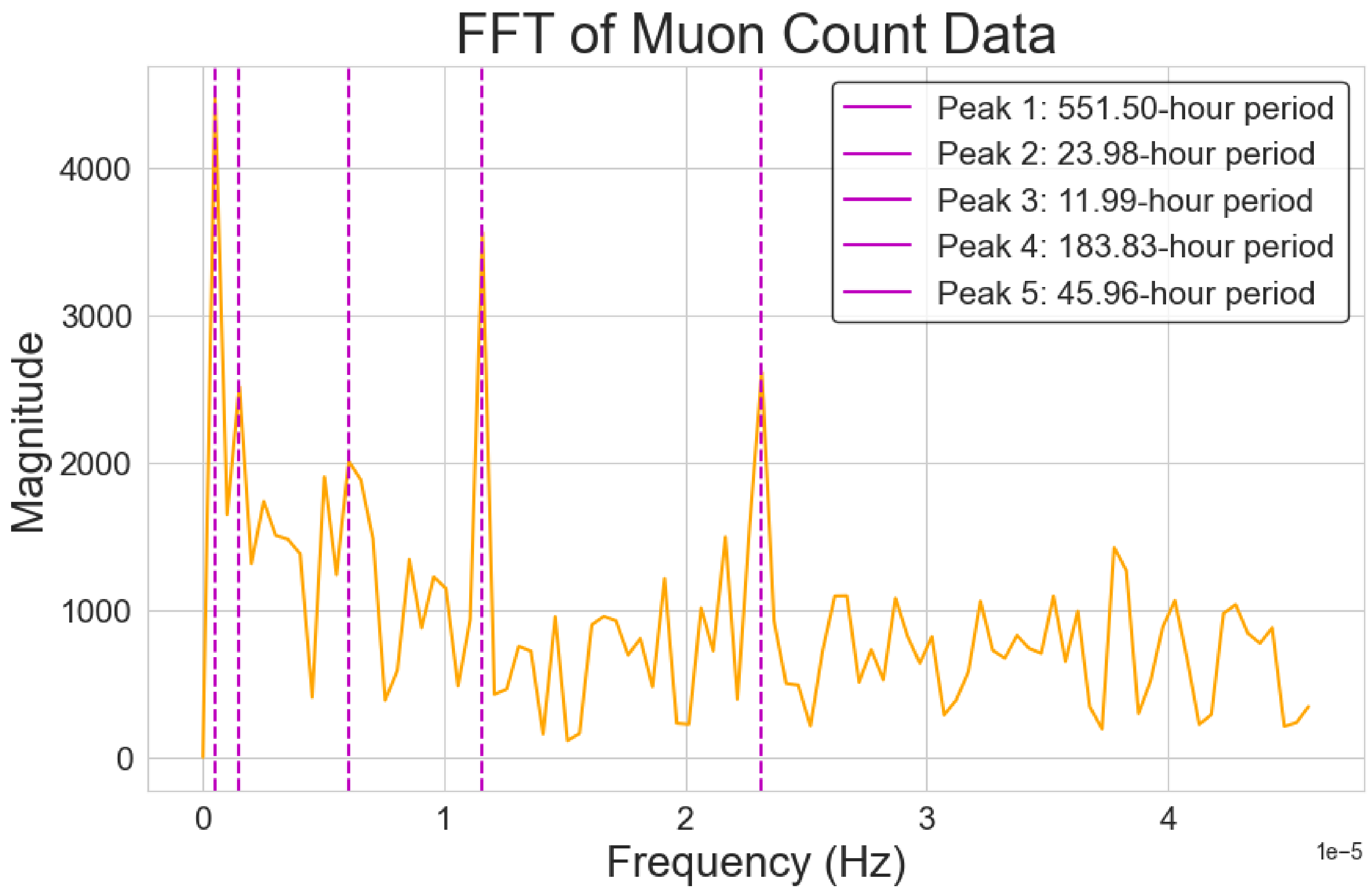
The mean and median are to indicate any skew, but we see the data approximates a Poisson process. Red lines indicate outliers determined by 1.5*IQR. When binned by hour, counts are not strictly a Poisson variable.

Full FFT (dataset 1)



Full FFT for dataset 1: 02/27 – 03/20. Smallest resolvable component has period 2*sampling interval = 2 hours. Data must be detrended (zero out the average or background noise) to perform an FFT, otherwise it blows up at 0 Hz.

What if 5-minute counts?



A finer sampling interval of 5-minutes (makes our data more Poisson-ian) improves the focus on the periodicities. Too small though, and the Poisson noise becomes statistically significant relative to the signal.

Unlucky equinox data

March 19, 2025

*“**EQUINOX GEOMAGNETIC STORM WATCH:** Minor G1-class geomagnetic storms are possible on March 20th when a CME might graze Earth's magnetic field... At this time of year, even a glancing blow can cause a geomagnetic storm thanks to the equinox Russell-McPherron effect... the **Russell-McPherron effect** happens around the equinoxes when magnetic fields on the sun and Earth can align and link up. This creates a sort of superhighway for solar wind to flow directly into Earth's magnetosphere ”*

The equinox and coinciding increased geomagnetic storm activity might’ve been reflected in our data if it did not cutoff on March 20.

Just blame everything on GM storms

March 26, 2025
*"**GEOMAGNETIC STORMS ARE UNDERWAY**: Geomagnetic storms are **underway on March 26th** as Earth enters a stream of fast-moving solar wind flowing from a giant hole in the sun's atmosphere... Earth should remain inside the stream for the next 24 hours..."*

April 9, 2025
*"**GEOMAGNETIC STORM WATCH (G1)**: ...storms are **possible on April 11th** when a CME is expected to graze Earth's magnetic field. The CME was hurled into space by an erupting solar filament on April 8th."*

April 14, 2025
*"**GEOMAGNETIC STORM WATCH (G2)**: Geomagnetic storms are **likely on April 16th** when a pair of closely-spaced CMEs is expected to hit Earth's magnetic field."*

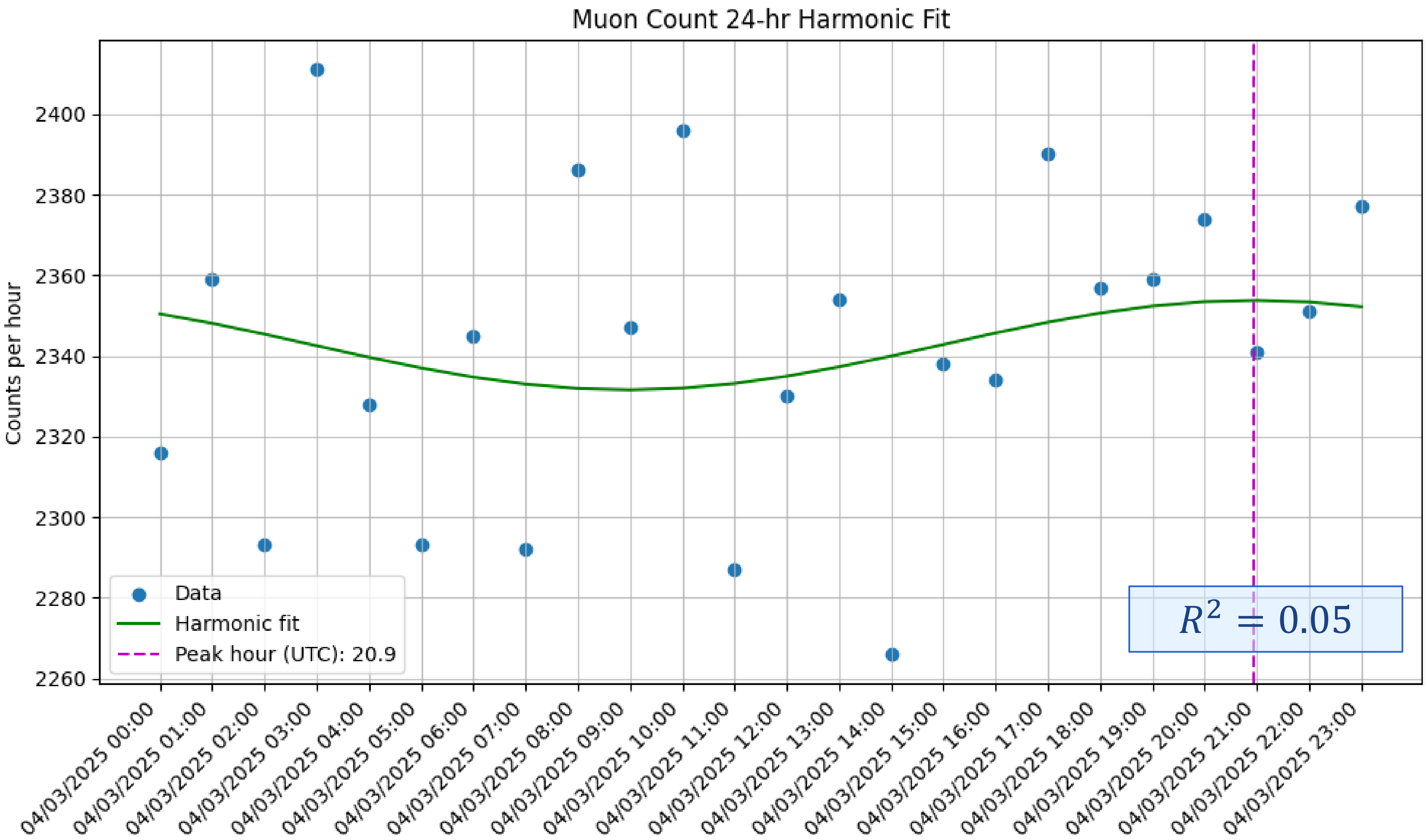
It's possible we can loosely correlate the hotspots in count activity to these geomagnetic storms. But positive spikes might be from weather: low pressure systems.

Correlation stats

	Size	P_{avg}	P Ci	S_{avg}	S Ci
IMF	27	-0.06 ±0.18	-0.40—0.32	0.06 ±0.21	-0.34—0.47
SWS	39	0.21 ±0.13	-0.08—0.44	0.16 ±0.15	-0.15—0.45
R SNo.	26	0.26 ±0.18	-0.16—0.59	0.18 ±0.21	-0.29—0.54
F10.7	41	0.29 ±0.13	0.03—0.52	0.28 ±0.14	-0.03—0.54

There is generally higher error and lower confidence for the smaller datasets. *Bootstrapping* resamples our small dataset and recalculates the coeffs 1000 times to give us estimates on these stats.

why do i try

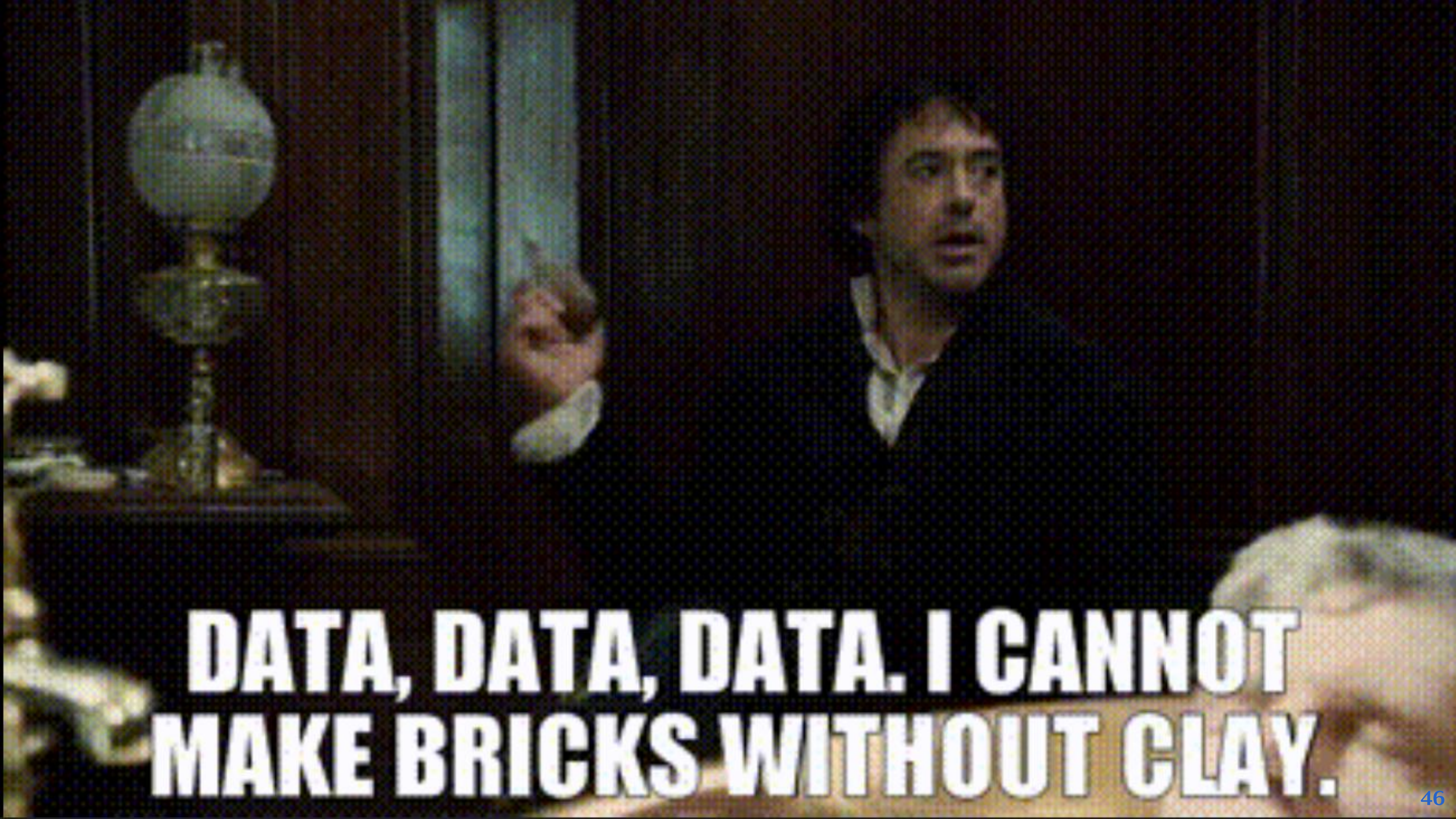


Many fits are clearly just capturing noise. The average R^2 was 0.1956 (like really bad). But hopefully the results still cluster around meaningful values in the aggregate.

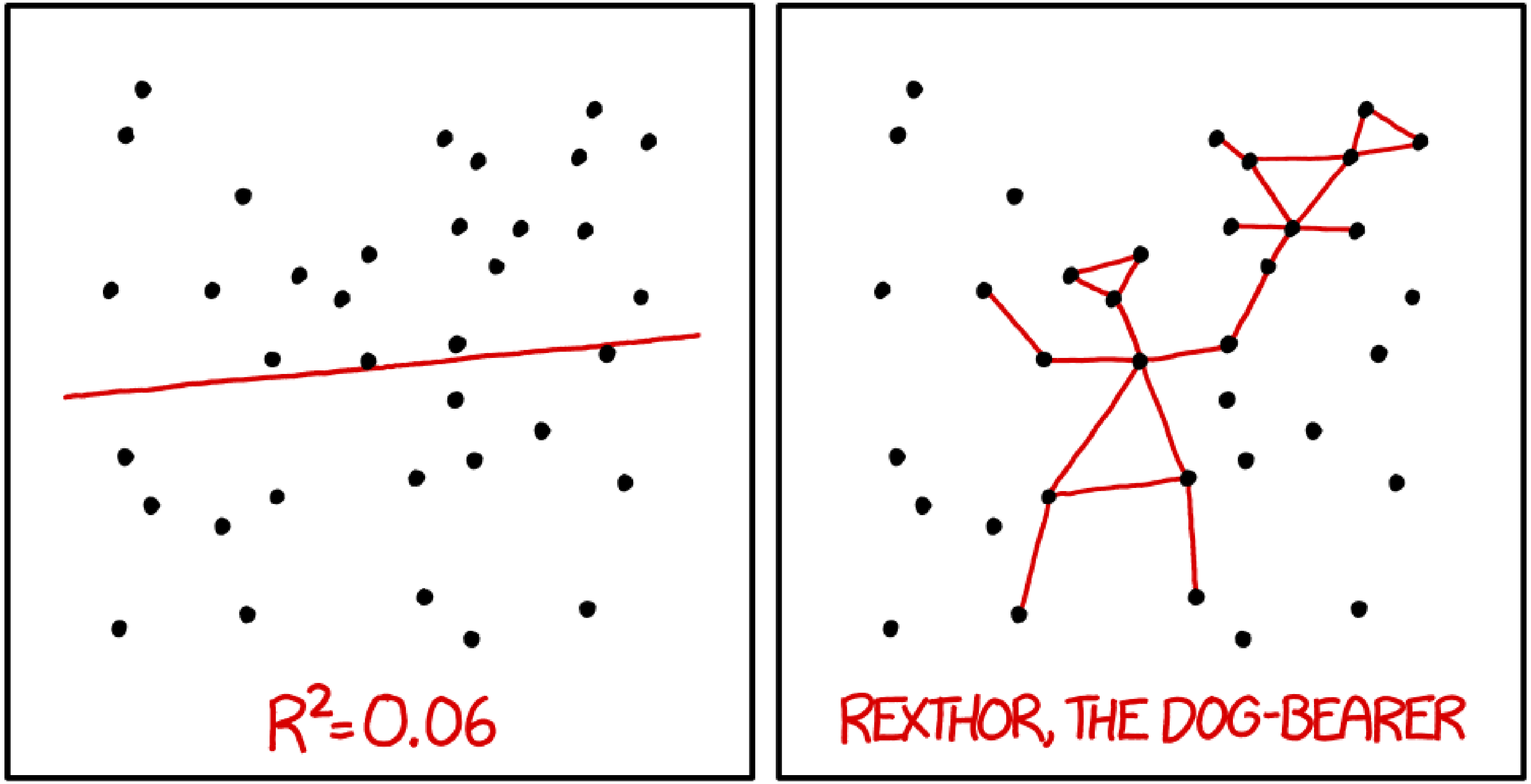
Correlating MC amplitude one day later

	P_{orig}	$P_{shifted}$	S_{orig}	$S_{shifted}$
IMF	-0.08 $\pm$0.19	0.18 $\pm$0.26	-0.11 $\pm$0.21	0.08 $\pm$0.21
SWS	0.49 $\pm$0.13	0.15 $\pm$0.16	0.44 $\pm$0.13	0.15 $\pm$0.18
R SNo.	-0.29 $\pm$0.18	-0.32 $\pm$0.21	-0.33 $\pm$0.20	-0.30 $\pm$0.20
F10.7	-0.12 $\pm$0.15	-0.12 $\pm$0.15	-0.15 $\pm$0.16	-0.16 $\pm$0.17

Similar results for the amplitude. Yeah. Science is hard.


DATA, DATA, DATA. I CANNOT
MAKE BRICKS WITHOUT CLAY.

xkcd 1725



I DON'T TRUST LINEAR REGRESSIONS WHEN IT'S HARDER TO GUESS THE DIRECTION OF THE CORRELATION FROM THE SCATTER PLOT THAN TO FIND NEW CONSTELLATIONS ON IT.