Exploring Correlations Between Recorded Weather Data, Instrument Data, and Resultant Noise Properties

Julián Ramos

August 2021

Abstract

We want to relate the noise in MUSTANG-2 maps to recorded atmospheric variables. We used data from scans taken in the past 4 years with MUSTANG-2. Each scan contains weather data for: zenith opacity, pyrgeometer, cloud coverage, air temperature, relative humidity, Sun azimuth, Sun elevation, windchill, wind-speed and wind-direction. For instrument data RMS was used as the dependent variable and any given weather variable as the independent one. A linear regression of the RMS was created using all the variables and their coefficients. This regression was then correlated against any given variable and its RMS. A correlation of 0.605 was measured between the RMS values and the regression. Correlations between the regression and noise levels in the maps of the different variables were sparse to moderate ($r \leq 0.371$). A specific correlation between the Weather and the Noise levels from any atmospheric variable could not be established. Further investigation should be done taking into account more Instrument data.

1 Introduction

The MUSTANG-2 receiver is a bolometer camera composed of 223 feed horns. It operates within the 75GHz and 105GHz frequencies with a 30 GHz bandwidth. Its operates on the Green Bank Telescope (GBT) in the National Radio Quiet Zone (NRQZ). MUSTANG-2 takes data as function of time, after the data is processed a map is created. The principal thing that affects the maps is noise. In a noisy map the flux of a faint object could get buried under the noise. Currently its unknown how the weather could affect the noise levels during MUSTANG-2 observations or what the Weather noise properties might be at this frequencies.

From the early days of ground-based millimeter astronomy it has been known that molecular oxygen (O2) and water (H_2O) vapor are the primary modulators of opacity (e.g. Kutner, 1978). If the atmosphere were perfectly uniform then it would not impart "noise" as much as a small signal which can be removed by scanning on and off source. However, the atmosphere's structure is turbulent; that is, it is changing with time and has differing fluctuations at different spatial (and therefore angular) scales (see e.g. Sayers et al., 2010). Because atmospheric emission will scale roughly with opacity, we should expect opacity to be a primary atmospheric contribution to noise (e.g. Adam et al., 2014).

In this project, we do not assess noise contributions in the "raw" data: the timestreams or Time Ordered Data (TODs). Rather, given an established pipeline (MIDAS, Romero et al., 2020), we choose to investigate how weather may affect the resultant noise in our maps (per scan). These scans, which span four years of observing, were reduced with a uniform set of parameters. With this research we aim to find correlations between the weather variables and the map noise. We expect that this work will help future observers anticipate how noisy their data will be based on instantaneously available weather information (data).

2 Method

We used scans taken over 4 years of observing with MUSTANG-2. The MIDAS infrastructure makes accessing certain variables such as mean wind speed, max wind speed, modified Julian Date (MJD), and telescope elevation very easy. Additional weather values such as zenith opacity (τ) were obtained by using CLEO to estimate the opacity at 90 GHz. Other values yet were obtained from weather log files. Thus, each scan contains values for: zenith opacity (τ), wind speed, wind direction, air temperature, humidity, dew point, cloud coverage, pyrgeometer, Sun azimuth and elevation. With these values we can calculate the air-mass (Equation 1), total opacity (Equation 2), easterly wind-speed component (Equation 3) and northerly wind-speed components (Equation 4).

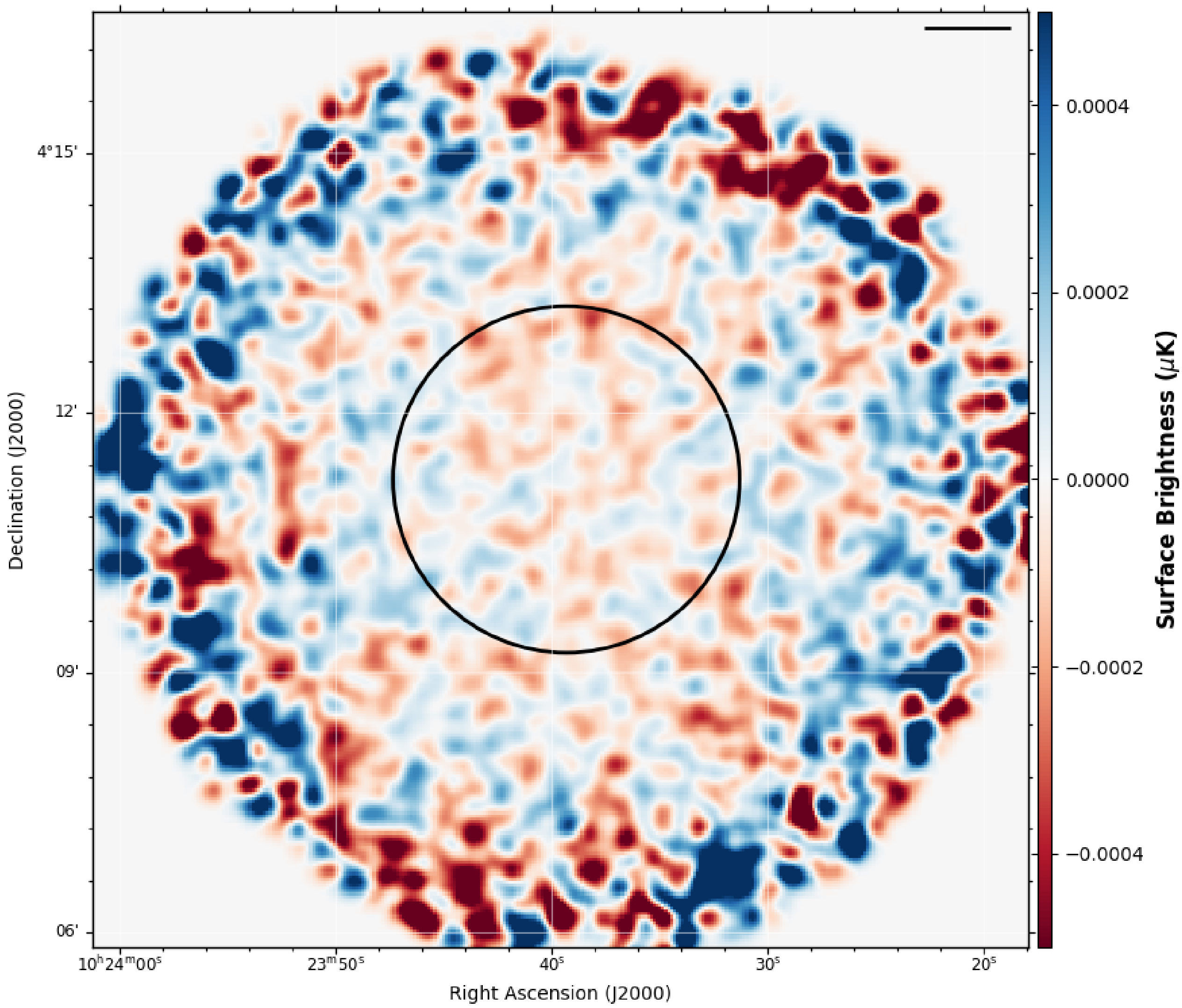

Figure 1: Here is a noise realization of a single scan. The circle indicates a radius of 2 arcminutes, within which the noise (RMS) is roughly uniform.

The map noise (per scan) is taken as the RMS about the central 2 arcminutes of the scan (Fig.1). This is the canonical quantification of noise for MUSTANG-2, scans in this region are fairly uniform. [1]. Regarding the specifics of the RMS calculation: each map has the original data (TOD) flipped so that it acts as a noise realization and the map is smoothed by a Gaussian of $10''$ full width-half maximum (FWHM) We made sure all of our data corresponds to scans with a radius of < 3 arcminutes. We need to take scan size into account in order to ensure all the data is within the same S/N ratio. Data with different S/N would lead us to inconsistent noise levels.

$$\text{airmass} = \frac{1}{sin(\text{elevation})} \tag{1}$$

$$\text{total opacity} = \frac{\tau}{sin(\text{elevation})} \tag{2}$$

$$\text{Northely wind speed} = cos(wind\,direction) \cdot (wind\,speed) \tag{3}$$

$$\text{Easterly wind speed} = sin(wind\,direction) \cdot (wind\,speed) \tag{4}$$

The data set containing the pyrgeometer, cloud coverage and effective temperature values only lasts for 1 year in the server after it is erased and replaced with blanking values. To keep our data consistent, scans with blanking values and outliers were discarded. From a total of 3222 scans over a 4 year period we used 836 scans.

The RMS was used as the dependent variables and weather variables where considered to be independent. A linear regression of the RMS was created using the weather variables and their coefficients. This regression was then correlated against any given weather variable and its corresponding RMS value to find a possible relationship. A fit using only two variables was used to compare the results with.

Correlations among the weather variables were found. Redundant variables were eliminated from the regression in order to improve it. Because the dew point correlated highly with temperature it was omitted from the regression. Earth IR as measured by the pyrgeometer and temperature were expected to have a correlation due to the green house effect. Both were kept in the regression. For the temperature there were multiple sources available from the different weather stations. All values from the stations correlated with the receiver temperature. Only one value was used.

Correlations between variables	
Weather Variables	Correlation (r)
dew point vs temperature	0.98
pyrgeometer vs zenith opacity	0.65
temperature vs zenith opacity	0.66

Table 1: Weather variables correlations. We used this information to remove redundant variables, in this case the dew point.

3 Results

When correlating the weather variables against their corresponding RMS in the scan the results were sparse to moderate ($r \leq 0.370$). The highest correlation amplitude was found with cloud coverage ($r = 0.370$), and the lowest one with the Easterly wind-speed component ($r = 0.013$). To improve the possibility of finding correlations a linear regression was created out of all the variables. To test our regression we plotted it against the RMS values that we want to predict. The higher the correlation is, the better the regression predicts the noise levels. We obtained a value of $r = 0.605$(Fig.2).

The regression with the simultaneous fit(Fig.3) gave us the same correlations amplitude as the variables against their corresponding RMS. The regression with the individual fit (Fig.4) used two variable and it also gave us the same correlations amplitudes as the simultaneous one. The slopes for the pyrgeometer, Sun azimuth, windchill and Northerly wind-speed changed in direction.

[1]See https://greenbankobservatory.org/science/gbt-observers/mustang-2/ for profiles of map noise (RMS) and scan sizes.

Correlations with RMS		
variable	correlation (r)	Regression Coefficient
zenith opacity	0.206	4.43×10^{-4}
pyrgeometer	0.288	-1.24×10^{-6}
cloud coverage	0.370	2.19×10^{-6}
air temperature	0.126	2.35×10^{-7}
humidity	-0.330	4.31×10^{-7}
Sun azimuth	-0.102	-3.63×10^{-7}
Sun Elevation	0.157	2.67×10^{-5}
airmass	0.039	3.46×10^{-5}
total opacity	0.191	2.72×10^{-4}
windchill	-0.357	-2.07×10^{-6}
wind-speed	0.269	1.43×10^{-6}
wind-direction	0.206	8.06×10^{-8}
wind-speed difference	0.340	2.08×10^{-5}
elevation	-0.068	4.23×10^{-7}
Northely wind-speed	0.031	-6.45×10^{-7}
Easterly wind-speed	-0.013	6.74×10^{7}

Table 2: Weather variables vs RMS Correlations (r) and coefficients used in the regression. All within 3 arcminutes.

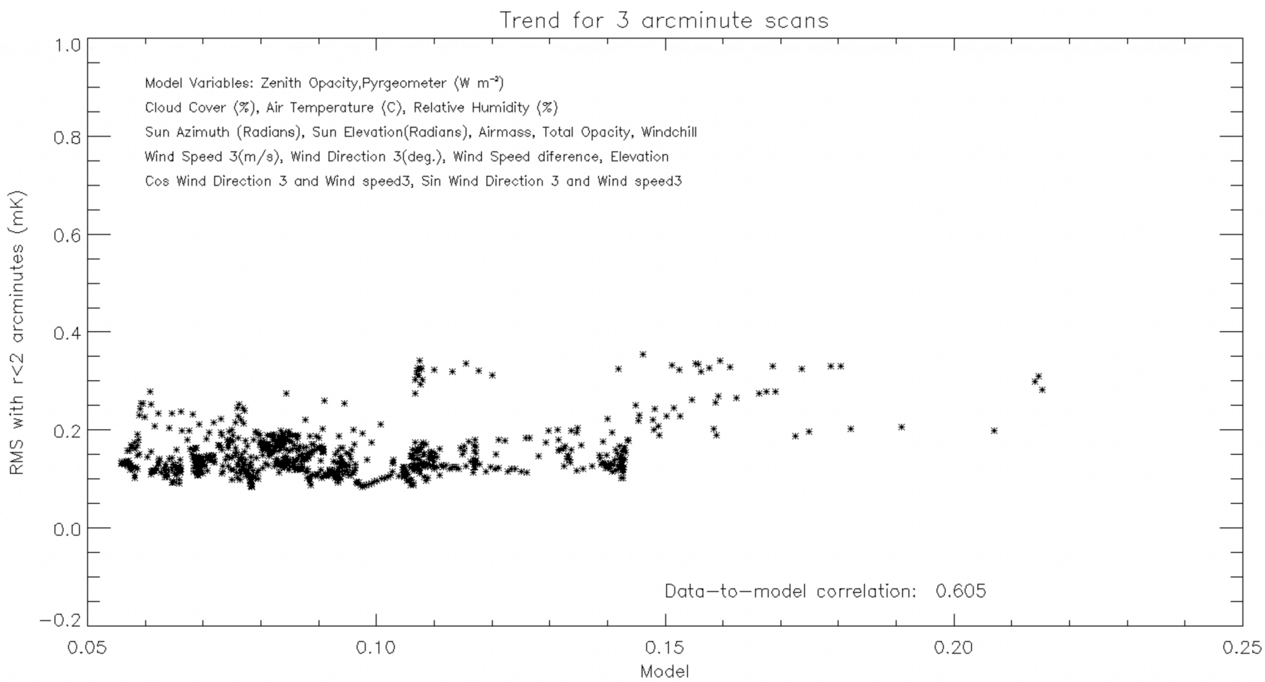

Figure 2: A correlation of 0.605 was established between the regression and the RMS values. The variables used were: zenith opacity, pyrgeometer, cloud coverage, air temperature, relative humidity, Sun azimuth, Sun elevation, airmass, total opacity, windchill, wind-speed and wind-direction.

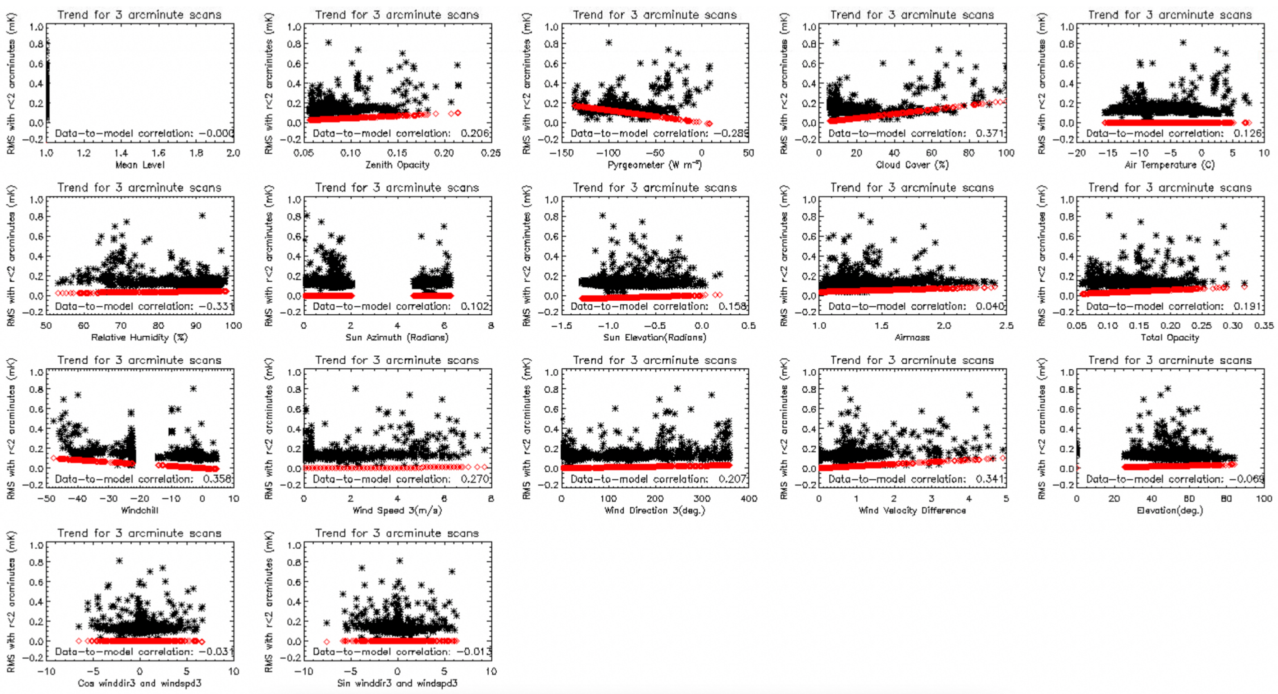

Figure 3: Simultaneous fit, all weather variables were used. In these graphs the regression (in red) was correlated against the weather variables vs RMS within 3 arcminutes.

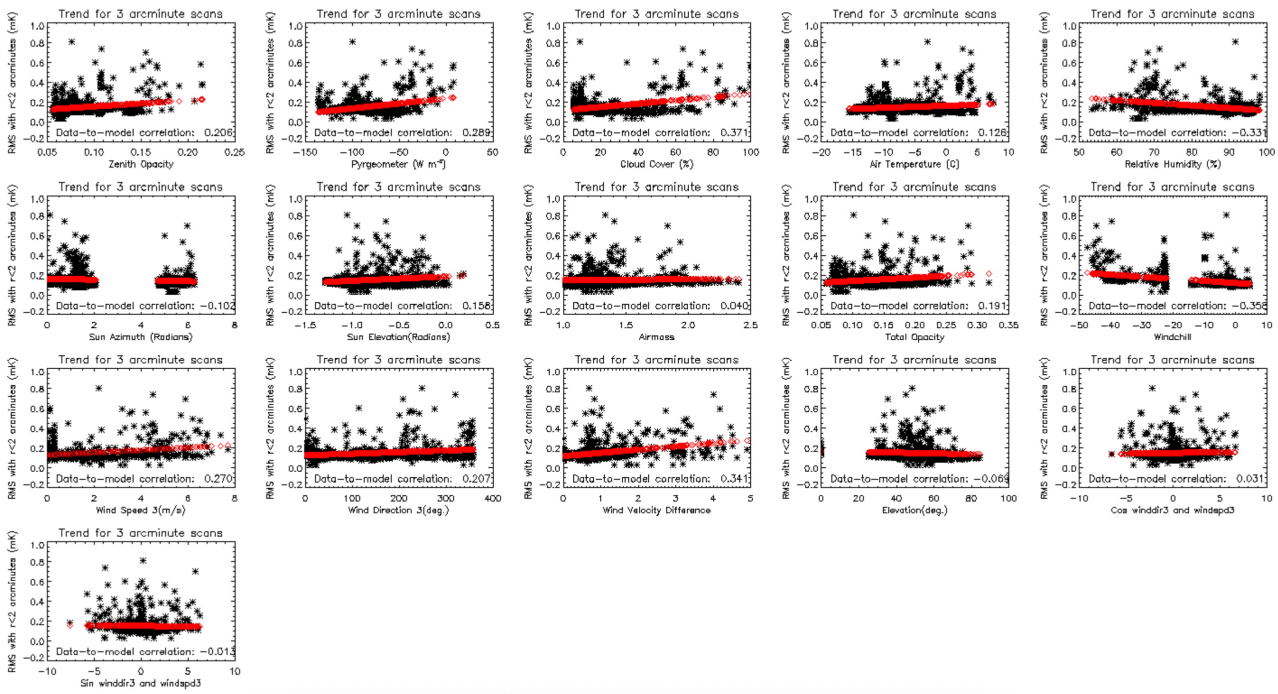

Figure 4: Individual fit, two ariables were used. In these graphs the regression (in red) was correlated against the weather variables vs RMS within 3 arcminutes.

4 Discussion

An unexpected correlation was found with the total opacity and zenith opacity. We expected the total opacity to have a higher correlation with the noise due to the larger area than just the zenith.

Correlations against the regression gave the same r value as the variables vs their RMS in the simultaneous and individual fit. This indicates the even when the regression correlated to $r = 0.606$ it predicts the noise levels accurately. We expected to have correlations with noise levels of an $r > 0.8$ to conclusively relate any given variable with noise in the map, this was not the case. The Change in the slope from the Simultaneous to individual fit is due to the variables containing negative values and the way the linear equations is used in the regression.

5 Conclusions and Future work

We did not obtain results that could conclusively relate the noise in the maps with the weather variables. However, our regression is a good approximation for the RMS. This indicate that the weather variables do have influence in the noise levels of maps. While we have reason to have predicted that opacity would have a dominant impact on the noise in our maps, our results would suggest that opacity is not such a dominant factor. Given that our linear regression only achieved a correlation of with $r = 0.606$, there is room for improved analyses in the future.

Future work could be divided in two avenues:

- Adding more variables to the linear regression, e.g.
  - Sector quadrant
  - array temperature
  - number of detectors used
- Investigating other characterizations of noise as in Sayers et al. (2010):
  - Look at power spectra of noise in maps
  - Look at power spectra of TODs (in the time domain)

References

Adam, R., Comis, B., Macías-Pérez, J. F., et al. 2014, , 569, A66

Kutner, M. L. 1978, , 19, 81

Romero, C. E., Sievers, J., Ghirardini, V., et al. 2020, , 891, 90

Sayers, J., Golwala, S. R., Ade, P. A. R., et al. 2010, , 708, 1674