Draft version September 6, 2022
Typeset using LaTeX default style in AASTeX63

Tracking ALMA System Temperature with Water Vapor Data at High Frequency

Hao He,[1] William R. F. Dent,[2] and Christine Wilson[1]

[1] *McMaster University*
1280 Main St W, Hamilton, ON L8S 4L8, CAN
[2] *Joint ALMA Observatory*
Alonso de Córdova 3107, Vitacura , Santiago, Chile

Submitted to PASP

ABSTRACT

As the world-leading submillimeter telescope, the ALMA observatory is now putting more focus on high-frequency observations at Band 7 – 10 (frequencies from 300 – 890 GHz). However, high-frequency observations often suffer from rapid variations in atmospheric opacity that directly affects the measured system temperature $T_{\rm sys}$. Current $T_{\rm sys}$ measurements use discrete atmosphere (ATM) calibrations done every few minutes, which usually involve $10 \sim 20$ ATM calibrations within a <2 hour observation at high frequencies. In order to obtain more accurate flux measurements and reduce the number of ATM calibrations, a new method to monitor $T_{\rm sys}$ continuously is proposed using existing data in the measurement set. In this work, we demonstrate the viability of using water vapor radiometer (WVR) data to track the $T_{\rm sys}$ continuously. We find a tight linear correlation between $T_{\rm sys}$ measured using the traditional method and $T_{\rm sys}$ extrapolated based on WVR data with scatter of 0.5% – 3%. Although the exact form of the linear relation varies among different data sets and spectral windows, we can use a small number of discrete $T_{\rm sys}$ measurements to fit the linear relation and use this heuristic relationship to derive $T_{\rm sys}$ every 10 seconds. We also test this correlation with atmospheric transmission (ATM) modeling and find a good agreement on results between modeled data and the real observation. We apply the semi-continuous $T_{\rm sys}$ method on a few data sets from Band 7 to Band 10 and compare the flux measured using these methods. We find the discrete and continuous $T_{\rm sys}$ methods give us consistent flux measurements with differences up to 5%. Furthermore, the calibrated fluxes from different subsets of the observation, or for different observations of the same target, are slightly more consistent with the continuous $T_{\rm sys}$ method. This method can help to save up to 10% of time for ALMA high-frequency observations.

1. INTRODUCTION

1.1. *Flux Calibration in ALMA*

Calibration is the process by which the astronomer converts electronic signals from the telescope into meaningful astronomical data. Accurate calibration is crucial for the Atacama Large Millimeter/submillimeter Array (ALMA), as millimeter and submillimeter wavelength radiation will be adversely affected by the atmosphere and the electronic signal path in a variety of ways, and the antennas will also be affected by the observing environment (Remjian et al. 2019, ALMA technical handbook, Chapter 10). One of the important calibration processes is the amplitude and flux calibration. The aim of this calibration is to convert the raw visibilities (and auto-correlations) from the correlator into brightness temperature or flux density by carefully tracking the instrumental and atmospheric variations and determining accurate conversion factors. Because of the large and rapidly varying opacity of water vapor, standard calibration procedures are less accurate at submillimeter wavelengths. For the flux calibration, a well defined scientific goal can be elucidated and set as the requirement. In numerous meetings and discussions, the scientific community

Corresponding author: Hao He
heh15@mcmaster.ca

originally made clear its desire to reach 1% flux density accuracy (e.g. Bachiller et al. 2003, report of the spring 2003 ASAC meeting), which means that we must be able to determine the overall flux density scale (and apply it to the visibilities and total power measurements) to 1% accuracy. In addition, the capability of achieving a dynamic range of 10000 or higher in ALMA images means that we must track the amplitude fluctuations to better than 1% (Yun et al. 1998). A later study by Moreno & Guilloteau (2002) showed that it is impractical to achieve 1% at submillimeter wavelengths, and so a requirement of 3% has been adopted for frequencies >300 GHz. The current achieved calibration accuracy for ALMA is 5% at the lower bands (100 GHz), 10% in mid-bands (200–400 GHz) and 20% in the higher bands (>400 GHz) (Remjian et al. 2019). This paper is part of the work being done to improve the overall flux calibration accuracy.

Currently there are two flux calibration strategies, the astronomical flux calibration and the direct instrumental amplitude calibration. The astronomical flux calibration uses an astronomical source with known flux and scales up the recorded amplitude based on that flux standard. This method requires the astronomical source to be bright and have stable flux. Currently planets are used as the primary flux standard while quasars are also used as an alternative flux standard due to their compact size and availability over the sky. However, the estimated accuracy of the planets' fluxes is only about 10% (Remjian et al. 2019, 10.4.7). Therefore, people are still searching for ideal flux calibrators with high accuracy, especially at high frequencies. In addition to the flux calibrator standard, the relative sky transmission on the calibrator and the science target, as well as the time variability of the transmissions will also affect the overall flux calibration. This is the standard method currently used by ALMA. On the other hand, for a stable system, one can directly translate the measured counts in total power into flux units using a direct instrumental amplitude calibration method. Both methods rely on accurate measurement of the sky opacity and tracking its variations during the observation. At millimeter wavelengths, the changes in atmospheric transparency will usually be very modest, under 1% over 10 minutes about 80% of the time. Since the same amount of water vapor results in much larger opacities in the submillimeter, the transparency fluctuations in the submillimeter over characteristic calibration time scales will be much larger, typically several percent during median stability conditions and sometimes > 10%.

For ALMA, both calibration methods require the precise measurement of the system temperature $T_{\rm sys}$ and complex gain G. $T_{\rm sys}$ represents the total thermal noise of the measurement. $T_{\rm sys}$ includes contributions from the sky, receiver, and system losses, with a large contribution coming from the sky temperature. Since ALMA is equipped with receivers of sufficiently low noise, the sky noise often dominates the total thermal noise. Therefore, it is necessary to track the changes in system temperatures caused by the fluctuations in the atmosphere. Current ALMA $T_{\rm sys}$ measurements use discrete atmosphere (ATM) calibrations done every few minutes with a cadence depending on the observing band. At low frequencies (< 300 GHz), we generally perform 2 or 3 $T_{\rm sys}$ measurements over a typical hour-long observation due to the assumed small variation in the atmosphere transmission. At high frequencies (> 300 GHz), due to the rapid opacity change in atmosphere, we generally perform $10 \sim 20$ ATM calibrations per hour. So the time overheads just due to ATM calibration can become quite significant – up to 20-30% at the highest bands (9 and 10, at 600–900 GHz). Moreover the variations of $T_{\rm sys}$ on timescales faster than the ATM calibration interval are not tracked with this discrete ATM calibration method. Therefore, one of the major goals in high-frequency flux calibration is to track $T_{\rm sys}$ more closely while also reducing the time spent on discrete ATM calibrations.

1.2. $T_{\rm sys}$ *Measurements in Flux Calibration*

The system temperature ($T_{\rm sys}$) is the fundamental parameter to determine the system sensitivity and the real flux of the source. Tsys includes various contributions, and can be written in a basic form as (adapted from Mangum 2017)

$$T_{\rm sys} = \frac{1}{\eta_{\rm eff} e^{-\tau_{\rm sky}}} \left(T_{\rm rx} + \eta_{\rm eff} T_{\rm sky} + (1 - \eta_{\rm eff}) \times T_{\rm amb}\right) \tag{1}$$

where

- $T_{\rm rx}$ is receiver temperature
- $T_{\rm sky}$ is sky temperature
- $T_{\rm amb}$ is ambient temperature where spillover is assumed to be terminated
- $\eta_{\rm eff}$ is the forward efficiency. This is equal to the fraction of the antenna power pattern that is contained within the main beam and is currently assumed to be 0.95

- $e^{-\tau_{\rm sky}}$ is the fractional transmission of the atmosphere, where $\tau_{\rm sky}$ is equal to the atmospheric opacity along the target's line of sight.

Note this equation is for single sideband (SB) and sideband separating (2SB) receivers, which are used for ALMA Band 3 – 8 observation. For this configuration, the image sideband gain is assumed negligibly small. The Band 9 and 10 receivers are using the double sideband configuration and hence $T_{\rm sys}$ are calculated differently (Mangum 2017, Eq. 6). $T_{\rm sky}$ and $T_{\rm amb}$ can be further expressed as

$$\begin{aligned} T_{\rm sky} &= T_{\rm atm}(1 - e^{-\tau_{\rm sky}}) \\ T_{\rm amb} &\approx T_{\rm atm} \end{aligned} \tag{2}$$

where $T_{\rm atm}$ is the representative atmosphere temperature. Note this is an approximation for now. Therefore, by combining equation 1 and 2, $T_{\rm sys}$ can be calculated as

$$\begin{aligned} T_{\rm sys} &= \frac{1}{\eta_{\rm eff} e^{-\tau_{\rm sky}}} \left[T_{\rm rx} + T_{\rm amb}(1 - \eta_{\rm eff} e^{-\tau_{\rm sky}})\right] \\ &\approx \frac{1}{e^{-\tau_{\rm sky}}} \left(T_{\rm rx} + T_{\rm sky}\right) \ (\eta_{\rm eff} \sim 1) \end{aligned} \tag{3}$$

This equation suggests that the key parameters to measure $T_{\rm sys}$ are $T_{\rm rx}$ and $T_{\rm sky}$ (as $\tau_{\rm sky}$ can be derived from $T_{\rm sky}$ using Eq. 2).

In ALMA flux calibration, the intensity of the observed source is directly proportional to $T_{\rm sys}$ by the following equation (e.g. Brogan 2018)

$$S_{\rm final} \sim S_0 \times \sqrt{T_{\rm sys}(i) T_{\rm sys}(j)} \times \Gamma \tag{4}$$

where S_0 and $S_{\rm final}$ are the fluxes measured before and after the flux calibration. Γ is the antenna efficiency factor to convert K to Jy and i and j represent the two antennas forming the baseline. Note that ALMA uses an antenna-based calibration method to simplify the calibration process. Nearly all of the changes to the visibility function (e.g. atmosphere, system noise, amplitude changes, delay changes) can be decomposed into the two complex antenna-based gain factors associated with any baseline. This approach reduces the number of gain correction terms for an N-element array from $N(N-1)/2$ baselines to N antennas. In this case, $T_{\rm sys}$ is associated with each antenna and $S_{\rm final}$ is associated with each baseline.

On the other hand, $T_{\rm sys}$ also determines the achieved rms noise of the observation. ALMA uses a modified form of the radiometer equation (Condon & Ransom 2016) to calculate the rms in kelvin from $T_{\rm sys}$, as shown below.

$$\text{rms} \approx \frac{c T_{\rm sys}}{\sqrt{\Delta\nu t_{\rm int}}} \tag{5}$$

where $\Delta\nu$ is the frequency bandwidth, $t_{\rm int}$ is the integration time of the observation and c includes the quantization and correlator efficiencies, and is typically 0.8-0.9 for ALMA. Therefore, higher $T_{\rm sys}$ means the data has a larger noise within one observation. In addition, for ALMA the weighting function used to combine visibility data is inversely proportional to $T_{\rm sys}$ as

$$\text{Weight} \propto \frac{1}{T_{\rm sys}(i) T_{\rm sys}(j)} \tag{6}$$

1.3. *Traditional Method to Measure $T_{\rm sys}$*

As noted above, $T_{\rm sys}$ is highly dependent on the sky opacity (eq. 2). ALMA antennas use a two-load system for $T_{\rm sys}$ measurement in Band 3 and higher, which is different from the one-load system used for other radio telescopes. The two-load system in theory can achieve a $T_{\rm sys}$ measurement accuracy of 1%, which is significantly better than that of a one-load system ("chopper wheel") of 5% (Yun et al. 1998). For ALMA, $T_{\rm sys}$ is obtained from an atmospheric calibration (ATM-cal) scan where a hot load, ambient load and sky are consecutively placed in front of the feed using an Amplitude Calibration Device (ACD). Typically this process takes 30-40 seconds, including antenna slew time and overheads. At frequencies below about 400 GHz, where the system temperatures are stable, an ATM-cal scan is made every 10 to 20 minutes. However, at higher frequencies, and wherever the opacity is large and more variable, every scan on the astronomical target will have an associated ATM-cal measurement (Remjian et al. 2019), implying a cadence

Table 1. Summary of different methods to track $T_{\rm sys}$

	WVR[1]	AC & SQLD[2]
Advantage	1. The data is continuously calibrated to measure $T_{\rm WVR}$[a]	1. Directly proportional to $T_{\rm sys}$ when $\tau_{\rm sky}$ is small
	2. Does not suffer from internal electronic gain drift[b]	2. Has the same frequency coverage as the science target
Disadvantage	1. Has different frequency coverage as the science target	1. The data is not calibrated.
	2. Directly tracks $T_{\rm sky}$ not $T_{\rm sys}$	2. Suffers from the electronic gain drift[b] or gain variations
		3. For AC, no linearity correction in FDM mode.

Columns: 1. Water vapor radiometer data. 2. Auto-correlation and square law detector data.
References: a. Hills et al. (2001). b. Payne et al. (2001).

of ATM calibration as fast as once every 2-3 minutes. From the two-load system, we can also measure $T_{\rm rx}$. In this case, the measured voltage can be expressed as (Mangum 2002)

$$\begin{aligned} V_{\rm hot} &= K\ (T_{\rm rx} + T_{\rm hot}) \\ V_{\rm amb} &= K\ (T_{\rm rx} + T_{\rm amb}) \\ V_{\rm sky} &= K\ (T_{\rm rx} + T_{\rm sky}) \end{aligned} \tag{7}$$

where K is the gain to convert the temperature to the measured voltage. Therefore, $T_{\rm rx}$ and $T_{\rm sky}$ can be expressed as

$$\begin{aligned} T_{\rm rx} &= \frac{T_{\rm hot}V_{\rm amb} - T_{\rm amb}V_{\rm hot}}{V_{\rm hot} - V_{\rm amb}} \\ &= \frac{T_{\rm hot} - Y_1 T_{\rm amb}}{Y_1 - 1} \\ T_{\rm sky} &= \frac{V_{\rm sky}T_{\rm amb} - (V_{\rm amb} - V_{\rm sky})T_{\rm rx}}{V_{\rm amb}} \\ &= Y_2 T_{\rm amb} - (1 - Y_2)T_{\rm rx} \end{aligned} \tag{8}$$

where $Y_1 \equiv V_{\rm hot}/V_{\rm amb}$ and $Y_2 \equiv V_{\rm sky}/V_{\rm amb}$. Unlike the atmosphere, $T_{\rm rx}$ is relatively constant throughout the observation. Measurements performed by ALMA show fluctuations of $T_{\rm rx}$ are generally smaller than 1% during normal sidereal tracking. With the measurement of $T_{\rm sky}$, we can further derive the optical depth based on Eq. 2 and calculate the $T_{\rm sys}$ based on Eq. 3. In summary, the expressions for the key quantities to measure $T_{\rm sys}$ are

$$\begin{aligned} T_{\rm rx} &= \frac{T_{\rm hot} - Y_1 T_{\rm amb}}{Y_1 - 1} \approx {\rm const} \\ T_{\rm sky} &= Y_2 T_{\rm amb} - (1 - Y_2)T_{\rm rx} \\ T_{\rm sys} &\approx \frac{1}{e^{-\tau_{\rm sky}}}\ (T_{\rm rx} + T_{\rm sky}) \end{aligned} \tag{9}$$

Therefore, during each ATM cal, we point the array to hot load, ambient load and sky to measure $T_{\rm rx}$ and $T_{\rm sky}$ and then calculate the $T_{\rm sys}$ at that time.

1.4. *Candidate Data to Track the Continuous* $T_{\rm sys}$

As mentioned above, the current method takes extra time to obtain a spot measurement of $T_{\rm sys}$ every few minutes. If we want to continuously track $T_{\rm sys}$, in theory there are 3 types of measurement data available from ALMA to achieve this goal: Water Vapor Radiometer (WVR) data, auto-correlation (AC) data or square law detector (SQLD) data. We will describe where these data arise, and the theory behind each method below. The advantages and disadvantages of each method are summarized in Table. 1.

The WVR data are used by ALMA to track the optical depth of the water vapor along the line of sight to each antenna, and hence are used to correct for the resulting effective pathlength and delay errors. The WVRs do Dicke switching and have internal calibrated loads, so the output from each WVR is the calibrated sky temperature ($T_{\rm WVR}$) at 4 frequencies (184.19, 185.25, 186.485 and 188.51 GHz respectively; Hills 2004) around the 183 GHz water line

taken every 1.152 seconds (Remjian et al. 2019, Section A.6). By comparing $T_{\rm WVR}$ with $T_{\rm amb}$, we can calculate the precipitable water vapour (PWV), which is proportional to the atmospheric opacity caused by the water absorption. Since $T_{\rm sky}$ at our observing frequency (Eq. 9) and $T_{\rm WVR}$ are tracking sky temperatures at different frequencies, we would expect

$$\begin{aligned} \tau_{\rm sky} &\propto \tau_{\rm 183GHz} \\ T_{\rm sky} &\propto T_{\rm WVR} \quad (\tau_{\rm sky}, \tau_{\rm 183GHz} << 1) \end{aligned} \tag{10}$$

where $\tau_{\rm sky}$ is the sky opacity at the observed frequencies and $\tau_{\rm 183GHz}$ is the optical depth for the water vapor line. If optical depth is small enough, we would expect the observed temperature is proportional to the optical depth and hence the proportion relation between the two optical depths holds also for the two measured temperatures. Since the major change in $T_{\rm sys}$ is caused by the variation in $T_{\rm sky}$, we would expect that $T_{\rm WVR}$ is tracking $T_{\rm sys}$. The major advantage of using the WVR data to trace $T_{\rm sys}$ is that the radiometer is constantly monitoring the sky and internally calibrating itself. Therefore, we can extrapolate $T_{\rm sys}$ throughout the entire observation based on the WVR data. Furthermore, since $T_{\rm WVR}$ is internally calibrating and tracking the sky variation, it does not suffer from the internal electronic drift or small changes in system gain, which can affect the measured values of an uncalibrated signal (see Table 1)

Alternatively, we would expect $T_{\rm sys}$ is tightly correlated with the total power signal received by each antenna. To be more precise, the total power signal should be directly proportional to $(T_{\rm rx} + T_{\rm sky})$, which can be used to calculate $T_{\rm sys}$ given the optical depth $\tau_{\rm sky}$ using Eq. 9 (see detailed discussion in Section 3). The total power signal received by each antenna is measured by a square-law detector built into the ALMA signal path, whose data is also recorded in the datasets. Additionally, the autocorrelation data recorded in the measurement set should also give us the total signal received by each antenna. We would expect

$$T_{\rm rx} + T_{\rm sky} \propto P_{\rm AC} \propto P_{\rm SQLD} \tag{11}$$

where $P_{\rm AC}$ and $P_{\rm SQLD}$ are the power of auto-correlation data and SQLD data read from the measurement set, respectively. If $\tau_{\rm sky}$ is small, we would expect direct proportionality between $T_{\rm sys}$ and the total power received which could help us derive continuous $T_{\rm sys}$. In addition, both AC and SQLD data cover the same frequency rangels as the actual observed science data so we do not need to assume atmosphere variation has the same effect on data at different wavelengths (the ratio proportionality in Eq. 10).

In Fig. 1, we plot the correlation between AC and WVR and SQLD data. We can see the AC and SQLD data follow a tighter linear correlation. These two types of data are expected to be equivalent and thus should follow a proportional correlation. We see a slight offset from the linear correlation, which is likely due to the lack of linearity correction for the 3-bit quantisation in the digital samplers of the ALMA Baseline correlator, as well as small DC offsets in the SQLD. On the other hand, we can see that the WVR and AC data do not follow the same proportional relation. This can be caused by various reasons summarized in Table 1. In particular, the distribution in the AC data at similar WVR levels on the left panel is indicative of slightly different system gains in different scans during the observation. This is thought to be partly due to the imperfect reproducibility of a stepped attenuator setting, which needs to be set and reset between ATM calibration scans. In this case, we need to compare the two types of data to explore which one is better in tracking $T_{\rm sys}$.

1.5. *Outline of This Paper*

In the following sections, we will explore how well different data track $T_{\rm sys}$ measurements. In Section 2, we explore the viability of using $T_{\rm WVR}$ to track $T_{\rm sys}$. We find that $T_{\rm WVR}$ has a tight linear correlation with $T_{\rm sys}$ when both values are normalized. In Section 3, we explored the viability to use AC or SQLD data to track $T_{\rm sys}$. What we find is that those two types of data do not work well in tracking $T_{\rm sys}$. In Section 4, we describe our new calibration method to use alternative $T_{\rm sys}$ derived from $T_{\rm WVR}$ and how it compares to the original discrete calibration method.

For our analysis, we use measurements sets from several projects in Bands 7, 8, 9 and 10. We also include two projects with multiple measurements sets from Band 7 and 9. The summary of the data we use is given in Table 2.

2. WVR DATA TO TRACK $T_{\rm sys}$

In this section, we examine how well $T_{\rm WVR}$ tracks $T_{\rm sys}$ and explore how the correlation is affected by various parameter choices. We mainly use the dataset Band8 (uid://A002/Xdb7ab7/X1880b) for illustration purposes. Examples from additional datasets are given in Appendix A.

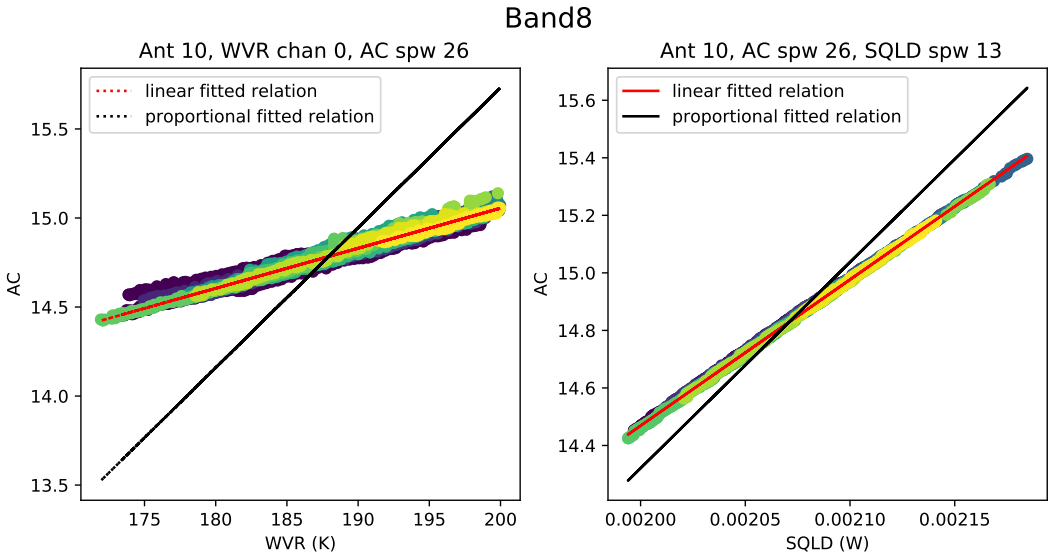

Figure 1. The correlation between water vapor radiometer (WVR) data, auto-correlation (AC) data and square law detector (SQLD) data for antenna 10 for dataset Band8 data. The scatter plots are color coded by the scan number. (Left) AC data versus WVR data. The red line is the best linear fit and black line is the proportional fit through 0. We have excluded WVR data greater than 200 as those are from samples taken on the hot load and ambient load. (Right) AC data versus SQLD data.

Table 2. Summary of Data

Dataset Label (1)	Project (2)	Band (3)	Target (4)	Data uid (5)	PWV (mm) (6)
Band10	2015.1.00271.S	10	Arp 220	uid://A002/Xbe0d4d/X12f5	0.28
Band9a	2016.1.00744.S	9	IRAS16293-B	uid://A002/Xbf792a/X14cc	0.37
Band8	2018.1.01778.S	8	SPT0311-58	uid://A002/Xdb7ab7/X1880b	0.85
Band7a	E2E8.1.00003.S	7	HT-Lup	uid://A002/Xec4ed2/X912	0.59
Band7b1	2018.1.01210.S	7	AS205A	uid://A002/Xda1250/X2387	0.69
Band7b2				uid://A002/Xda1250/X32df	0.53
Band7b3				uid://A002/Xda845c/X35d1	0.5
Band7b4				uid://A002/Xda1250/X3e39	0.51
Band7b5				uid://A002/Xda1250/X4db3	0.49
Band7b6				uid://A002/Xd99ff3/X15d7b	0.42
Band7b7				uid://A002/Xd99ff3/X1702c	0.51
Band7b8				uid://A002/Xd99ff3/X17da2	0.53
Band9b1	2019.1.00013.S	9	Circinus	uid://A002/Xed9025/X769c	0.43
Band9b2				uid://A002/Xed8123/X7b1	0.37
Band9b3				uid://A002/Xed4607/X1208a	0.34

Columns: (1) Label for each dataset (2) ALMA project code (3) Observed Band. (4) The target name to be observed (5) The uid for each measurement set (6) The precipitate water vapor (PWV).

2.1. $T_{\rm sys}$ *versus* $T_{\rm WVR}$

To check whether $T_{\rm sys}$ is tracked by the WVR data, we first need to match the WVR data taken at the same time as the $T_{\rm sys}$ measurements. We then average the WVR values that are within 10s around the time when $T_{\rm sys}$ is measured and compare the averaged $T_{\rm WVR}$ with its corresponding $T_{\rm sys}$. 10 s is a typical time for one ATM scan and hence is the

shortest timescale we expect $T_{\rm sys}$ to change. We also note that $T_{\rm sys}$ recorded in the measurement set is a spectrum with two polarizations. In our analyses to compare $T_{\rm sys}$ with $T_{\rm WVR}$, we average $T_{\rm sys}$ from both polarizations and from different spectral channels within one spectral window (spw).

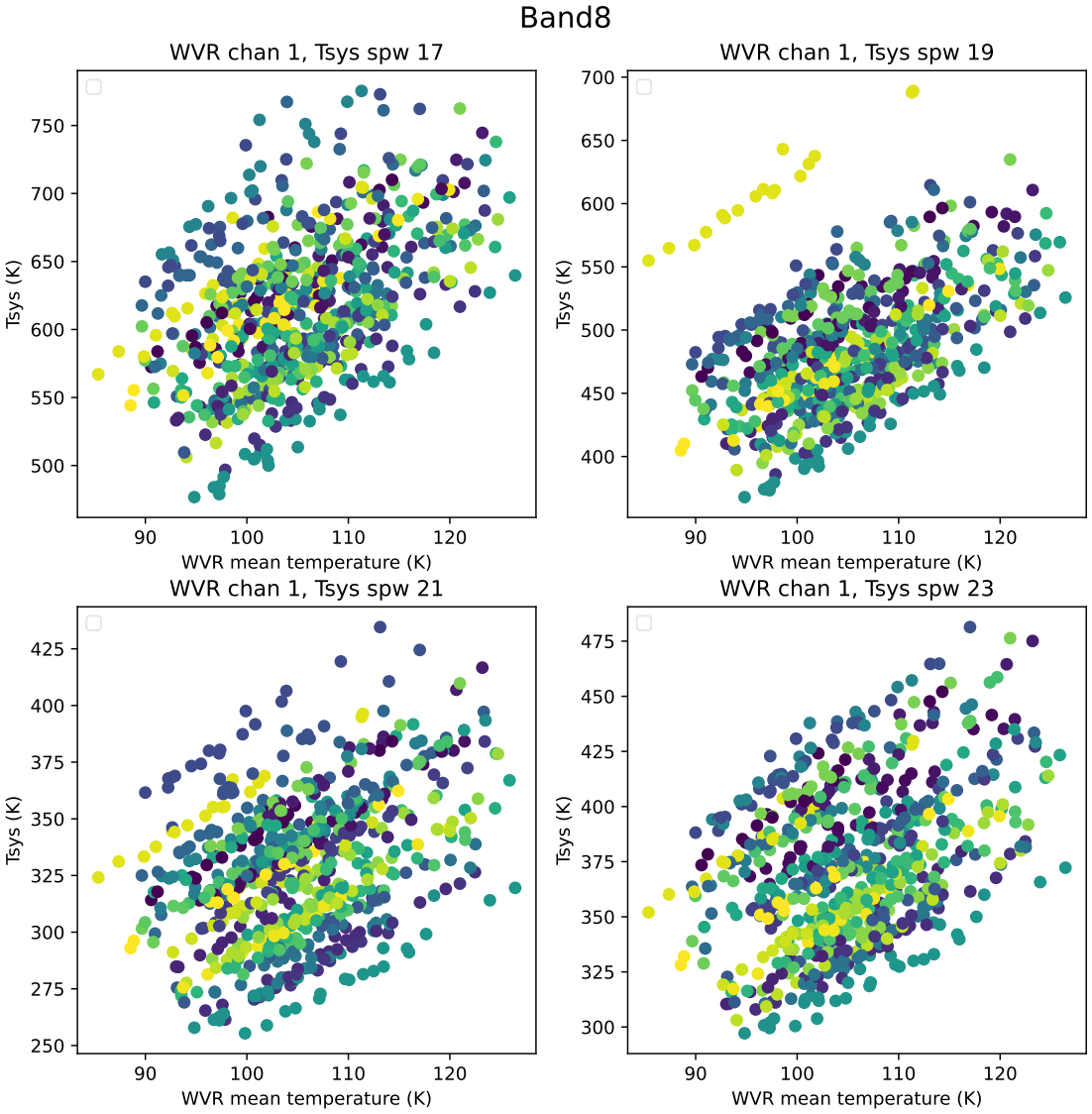

Figure 2. The correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ for different spws of all the antennas from dataset Band8. Different colors specify data points from different antennas. Note all these values are averaged among all spectral channels and both polarizations. See detailed description in Section 2.1.

We first plot $T_{\rm sys}$ versus $T_{\rm WVR}$ from all antennas for each data set. One example of $T_{\rm sys}$ versus $T_{\rm WVR}$ is shown in Fig. 2. For this case, we select $T_{\rm WVR}$ from WVR channel 1. We will discuss in Section 2.2 how the selection of different WVR channels affects the relation between $T_{\rm WVR}$ and $T_{\rm sys}$. As we can see, there is a significant correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ for each spw. However, there is a large scatter along the direction perpendicular to the trend. Due to the differences in the receiver ($T_{\rm rx}$ and sideband gains) and WVR between antennas, we would expect the same WVR value might correspond to slightly different $T_{\rm sys}$ values from different antennas. In Fig. 2, we can clearly see data points from a single antennae generally lie along the same straight line. Furthermore, since bandpass, phase-cal and science observations are observing targets at different elevations, it is possible that the scatter is also caused by

data from different types of observations. Therefore, to see if the WVR tracks the time variation of $T_{\rm sys}$, we normalize $T_{\rm sys}$ and $T_{\rm WVR}$ by the first measurement for each observing target (bandpass, phase-cal and science) of each antenna as

$$\begin{aligned} \hat{T}_{\rm sys,source}(t) &= \frac{T_{\rm sys,obs}(t)}{T_{\rm sys,obs}(1^{\rm st})} \\ \hat{T}_{\rm WVR,obs}(t) &= \frac{T_{\rm WVR,obs}(t)}{T_{\rm WVR,obs}(1^{\rm st})} \end{aligned} \tag{12}$$

where $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ is the normalized values of $T_{\rm sys}$ and $T_{\rm WVR}$, the subscript 'obs' is the generalized term for each type of observing target (bandpass, phase and science) and $i^{\rm th}$ in the bracket means the value when $i^{\rm th}$ $T_{\rm sys}$ for each observing target is measured. The correlation between the $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ is shown in Fig. 3. We can see that these two variables have a tight linear correlation, with scatter less than 1%. This tight linear correlation is also seen in other data sets, as illustrated in Appendix A. *This correlation indicates $T_{\rm WVR}$ can be used to track the $T_{\rm sys}$ variation.* However, we note that the fitted linear relation has offsets from the 1-to-1 proportional relation. We also see different spws seem to have fits with different slopes and intercepts. This is what we expect since $T_{\rm sky}$ instead of $T_{\rm sys}$ should be proportional to $T_{\rm WVR}$ according to Eq. 10. Therefore, the slopes and intercepts of the fits will be different due to differences in the atmospheric opacity at different frequencies (ie. different constants of proportionality in Eq. 10. Moreover the sky temperature includes contributions from opacity due to the wet component from H_2O lines (which are relatively wide due to pressure broadening), and from the dry component (mostly due to relatively narrow lines of O_3, but also including a continuum component as well as other molecules). $T_{\rm WVR}$ is dominated by the water line (wet) component, but $T_{\rm sky}$ around the observing frequency will depend on the relative wet and dry opacities averaged over the chosen spectral window. The slope and intercept of the correlation depends on details of the atmospheric transmission curves, and in Section 2.7 we compare with an atmospheric opacity model. However, for now, we just take the heuristic approach, to derive the constant from the data.

We can then use the fitted linear relation to extrapolate the continuous $T_{\rm sys}$ based on the first $T_{\rm sys}$ value for each observing target and the stream of $T_{\rm WVR}$ values. The exact equation can be expressed as

$$\begin{aligned} T_{\rm sys}(t) &= T_{\rm sys}(1^{\rm st}) \cdot \hat{T}_{\rm sys,obs}(t) \\ &= T_{\rm sys}(1^{\rm st}) \cdot \Big[m\ \hat{T}_{\rm WVR,obs}(t) + b \Big] \end{aligned} \tag{13}$$

where $T_{\rm sys}(1^{\rm st})$ and $T_{\rm WVR}(1^{\rm st})$ are $T_{\rm sys}$ and $T_{\rm WVR}$ values used to normalize each antenna and each type of observing targets. m and b are slope and intercept of the fitted linear function. For making the extrapolation, we also sample and average the WVR data every 10 seconds to be consistent with our fitting parameter choice. An example of extrapolated $T_{\rm sys}$ for one antenna is shown in Fig .4. As we can see, the extrapolated continuous $T_{\rm sys}$ is consistent with the original discrete $T_{\rm sys}$ values for all 4 spectral windows. The trend is also quite continuous with no obvious glitches due to the measurement noise. The trends for all 4 spectral windows are similar, which indicates $T_{\rm WVR}$ truly reflects the change in atmosphere.

Examples of fitting and $T_{\rm sys}$ extrapolation for other datasets are shown in Appendix A. We can see for all datasets, $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ have a tight linear correlation but with different slopes and intercepts. The extrapolation also works well for most of the data sets.

2.2. *Other WVR Channels*

The ALMA WVRs have 4 filter channels at frequencies (184.19, 185.25, 186.485 and 188.51 GHz respectively) close to the 183GHz H_2O line (Hills et al. 2001). These channels have different sensitivities to the line-of-sight water content (PWV) and hence Tsys, depending on the actual pwv at the observing time. In this section, we explore how the different parameter choices will affect the correlation between $T_{\rm sys}$ and WVR. In Section 2.1, we selected $T_{\rm WVR}$ in channel 1 to track $T_{\rm sys}$. Here we test using $T_{\rm WVR}$ from different WVR channels for $T_{\rm sys}$ on the same dataset (Fig. 5). As we can see, different WVR channels give us a similar tight correlation between normalized $T_{\rm sys}$ and $T_{\rm WVR}$. As expected, $T_{\rm WVR}$ from the different channels also gives us different slopes and intercepts. Therefore, if we average $T_{\rm WVR}$ from different channels, it should not reduce the scatter around the fitted linear function.

In our later analysis to apply the continuous $T_{\rm sys}$ in data calibration, we select WVR channels to maximize the following weighting function

$$w = \overline{T}_{\rm WVR}(\overline{T}_{\rm WVR} - 275) \tag{14}$$

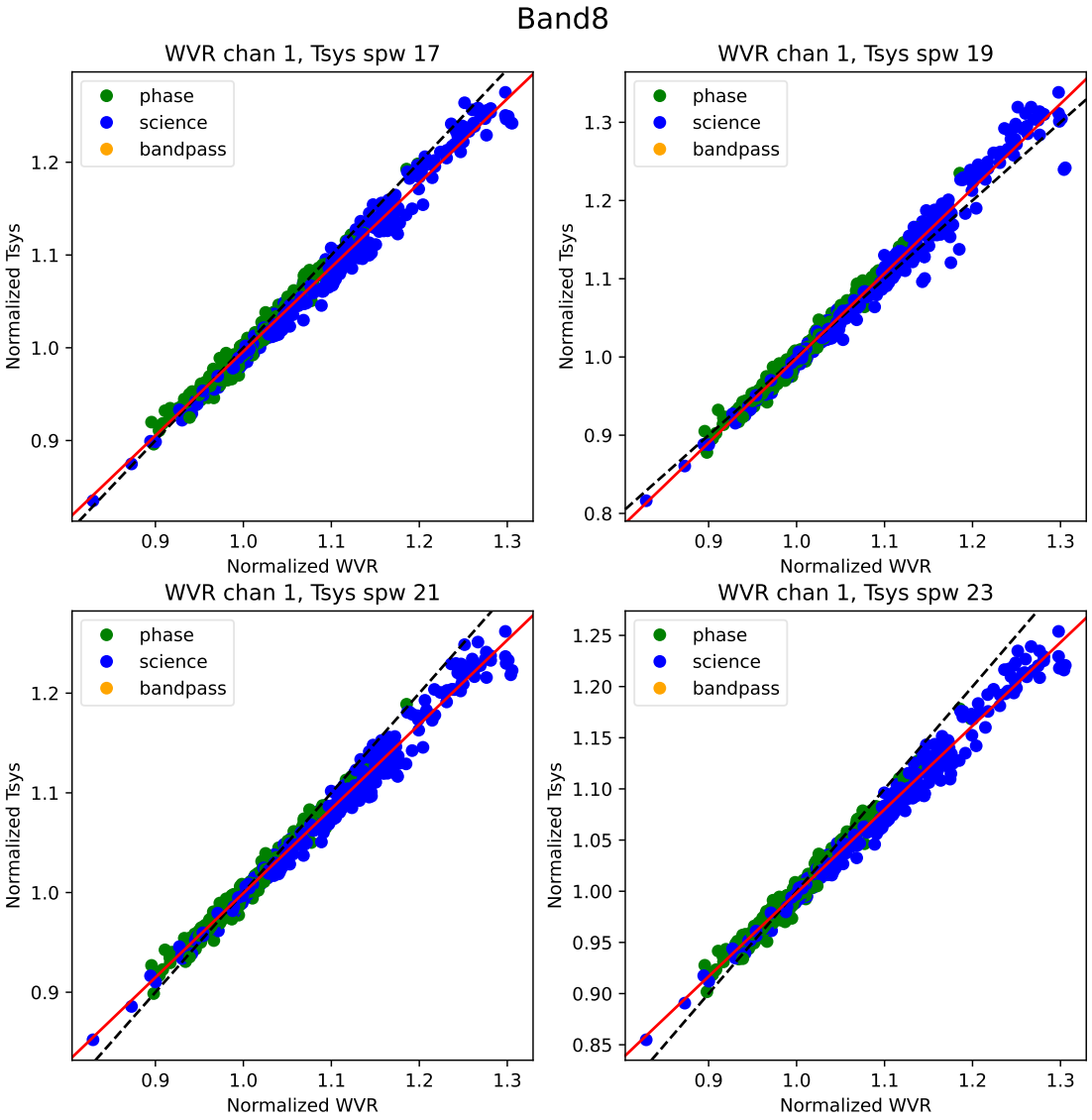

Figure 3. The same as Fig. 2, except now $T_{\rm sys}$ and $T_{\rm WVR}$ are normalized by the first measurement of each type observation (bandpass, phase and science). The orange, blue and green points are from bandpass, phase and science scans respectively. The red line is the linear fitting of the relation. The black dashed line indicates the 1-to-1 relation. See detailed description in Section 2.1.

where $\overline{T}_{\rm WVR}$ is the averaged $T_{\rm WVR}$ in one channel and 275 K is the approximate atmosphere temperature. The principle for this selection criterion is to make $T_{\rm WVR}$ neither too small to be robust against noise (in the case of low opacity) nor too large to be saturated (in the case of high opacity). Based on this criterion, we generally select WVR channel 0 or 1 for datasets in our analysis.

2.3. *Using PWV to track the* $T_{\rm sys}$

For single dish telescopes such as APEX and JCMT, WVR data has been used to continuously track optical depth at the observed frequencies (Dempsey et al. 2013). The method converts $T_{\rm WVR}$ values from 4 WVR channels into a single PWV value and use it to track the optical depth at any given time, which reduces the effect of measurement noise from a single channel. Given what we are doing is similar, as $T_{\rm sys}$ is mostly affected by the change in atmospheric optical depth, we can try to use PWV instead of $T_{\rm WVR}$ from a specific channel to track $T_{\rm sys}$. We calculate the PWV by fitting the Lorentz profile for the water line given the $T_{\rm WVR}$ from the 4 channels. We then normalize the PWV values

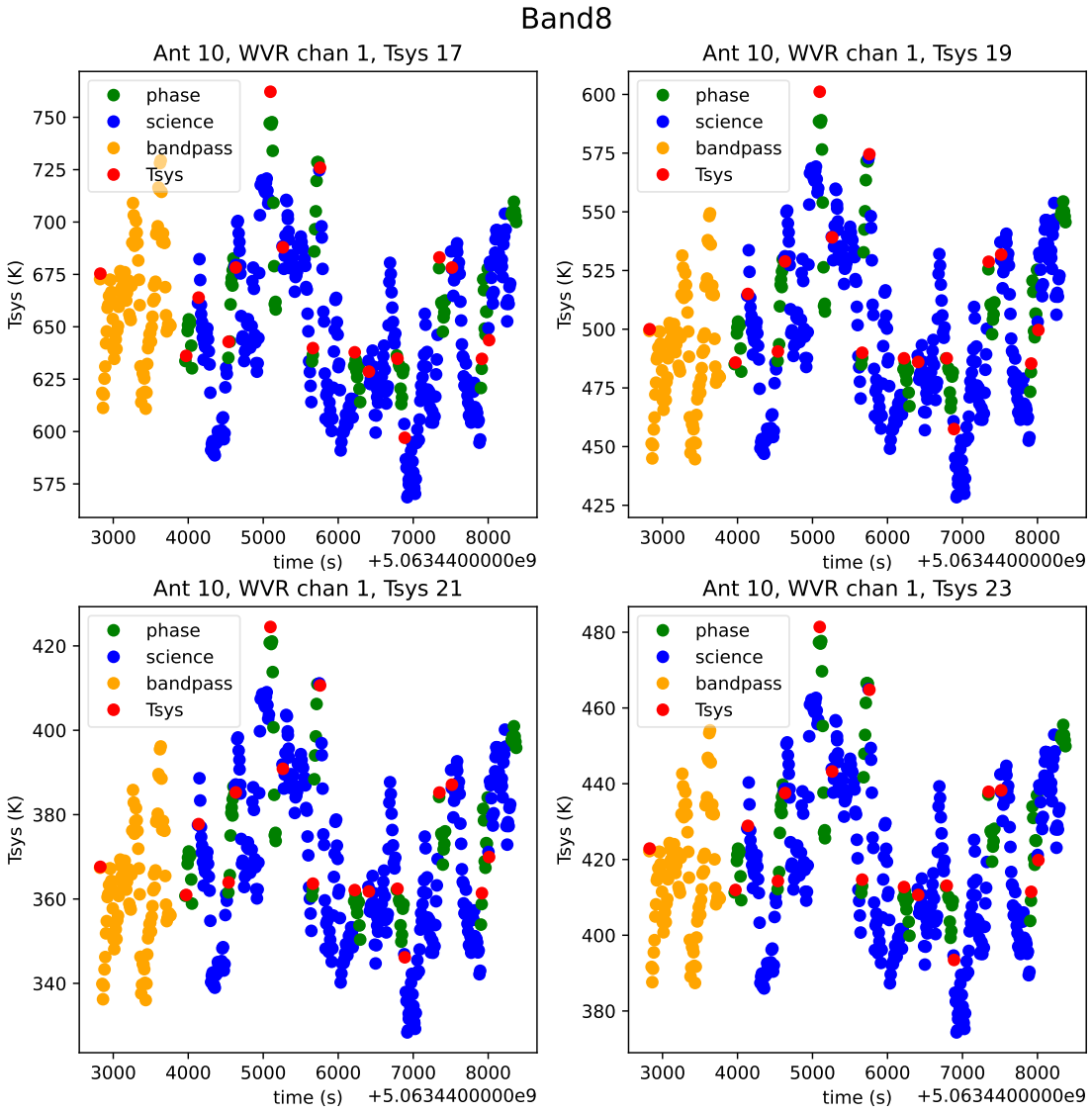

Figure 4. The extrapolated $T_{\rm sys}$ versus the original $T_{\rm sys}$ for antenna 10 of the same dataset used in Fig. 2. The orange, green and blue points are extrapolated continuous $T_{\rm sys}$ for each observing target based on Eq. 13 while the red points are original $T_{\rm sys}$ measurements.

the same way as we do for $T_{\rm WVR}$ (Eq. 12). The comparison between normalized $T_{\rm sys}$ and PWV values is shown in Fig. 6. We can see that PWVs also have a tight correlation with $T_{\rm sys}$ as expected. Compared with $T_{\rm WVR}$ in channel 2 and 3 (Fig. 5), the correlation between $T_{\rm sys}$ and PWV shows fewer outliers, which suggests we can also use PWVs to track $T_{\rm sys}$ to minimize the noise in $T_{\rm WVR}$ measurements. On the other hand, we still see different slope and intercepts between $T_{\rm sys}$ and PWV for different $T_{\rm sys}$ spws, which is also consistent with what we expect as different spws cover different frequency ranges and hence have slightly different portions of contribution from wet and dry components.

2.4. *Fewer ATM scans to Fit the Relation*

As discussed in Section 2.1, the $T_{\rm sys}$ in different spectral windows have different linear relations with $T_{\rm WVR}$. Therefore, we are still not able to predict the exact form of the linear relation before the data is taken. To achieve the goal to reduce the number of $T_{\rm sys}$ measurements, we will test in this section if we can get the same linear relation between $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ with fewer ATM scans. Since we need to calculate the normalized $T_{\rm sys}$, we need at least two ATM scans to fit the linear correlation. To make the fitting more robust, we use 4 ATM scans for the fitting with 2 from

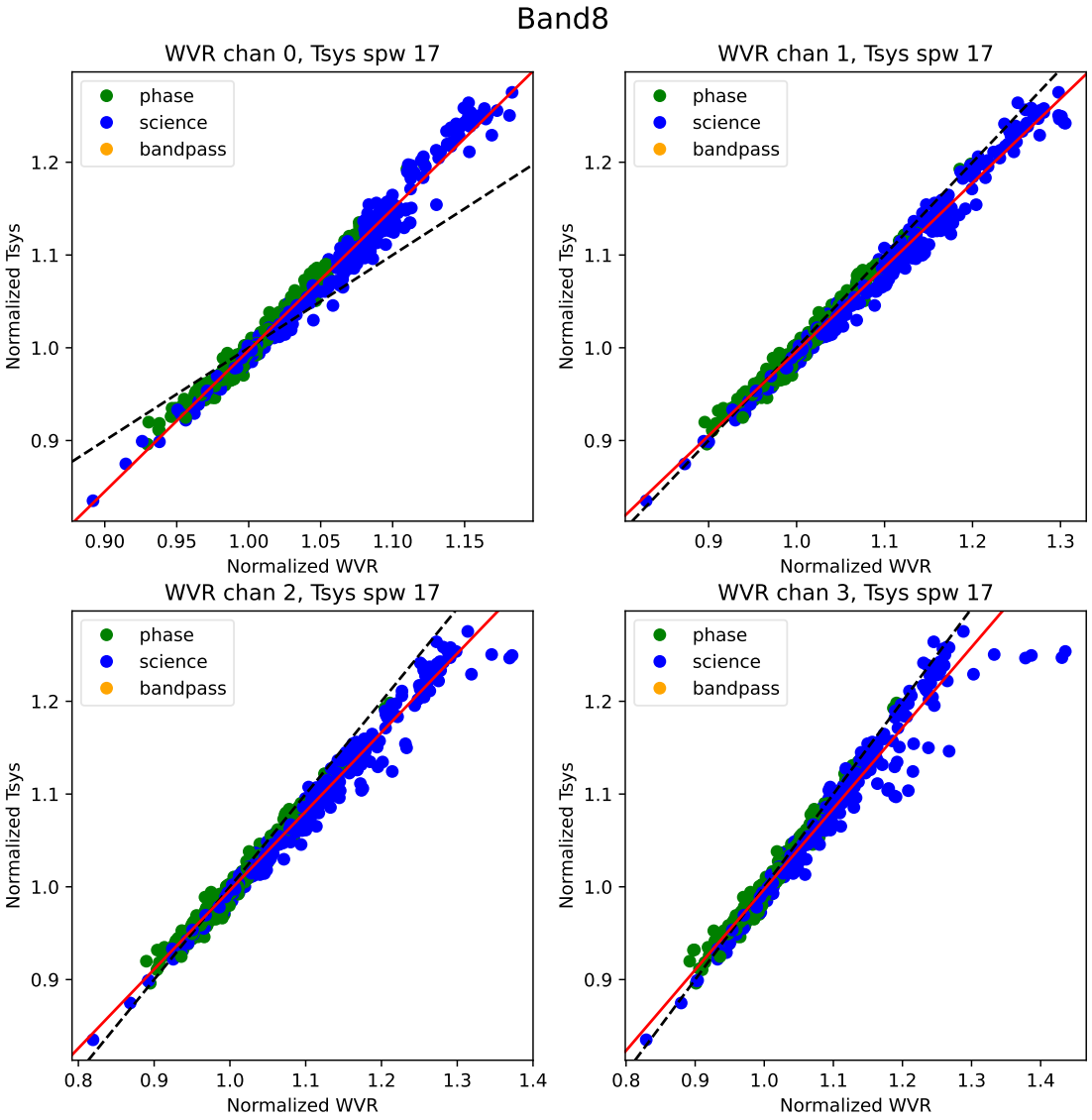

Figure 5. Correlation between the Normalized $T_{\rm sys}$ and $T_{\rm WVR}$ for different WVR channels. For $T_{\rm sys}$ we select the data from spectral window 17.

phase target and 2 from the science target. The 4 ATM scans give us 2 independent $\hat{T}_{\rm sys}$ values if we normalize the $T_{\rm sys}$ from phase and science target independently. For the ATM scan selection, we select 2 ATM scans at the start and 2 ATM scans in the middle. One example of the fits using 4 ATM scans is shown in Fig. 7. As we can see, the fits based on data from all ATM scans have almost no difference from the fits based on only 4 ATM scans. The scatter of all the data points around the new relation is $\sim 0.6\%$, which has almost no difference from the old scatter of $\sim 0.5\%$.

Fits using 4 ATM scans for other data sets are also shown in Appendix A. We can see that the fits do not change much for almost all data sets except for that in dataset Band9b. We will discuss the data in 2019.1.00013.S in Section 2.6. In Fig. 8, we plot the relative scatter around the fit versus the maximal difference divided by mean value of the $T_{\rm sys}$ for every $T_{\rm sys}$ spw of all the data sets. As we can see, the scatter using both fitting methods is generally below 3%. Besides, fitting with just 4 scans only slightly increases the scatter compared with fitting with all scans. From this quantitative comparison, we can see it is viable to reduce the number of discrete $T_{\rm sys}$ measurements when using $T_{\rm WVR}$ to track $T_{\rm sys}$.

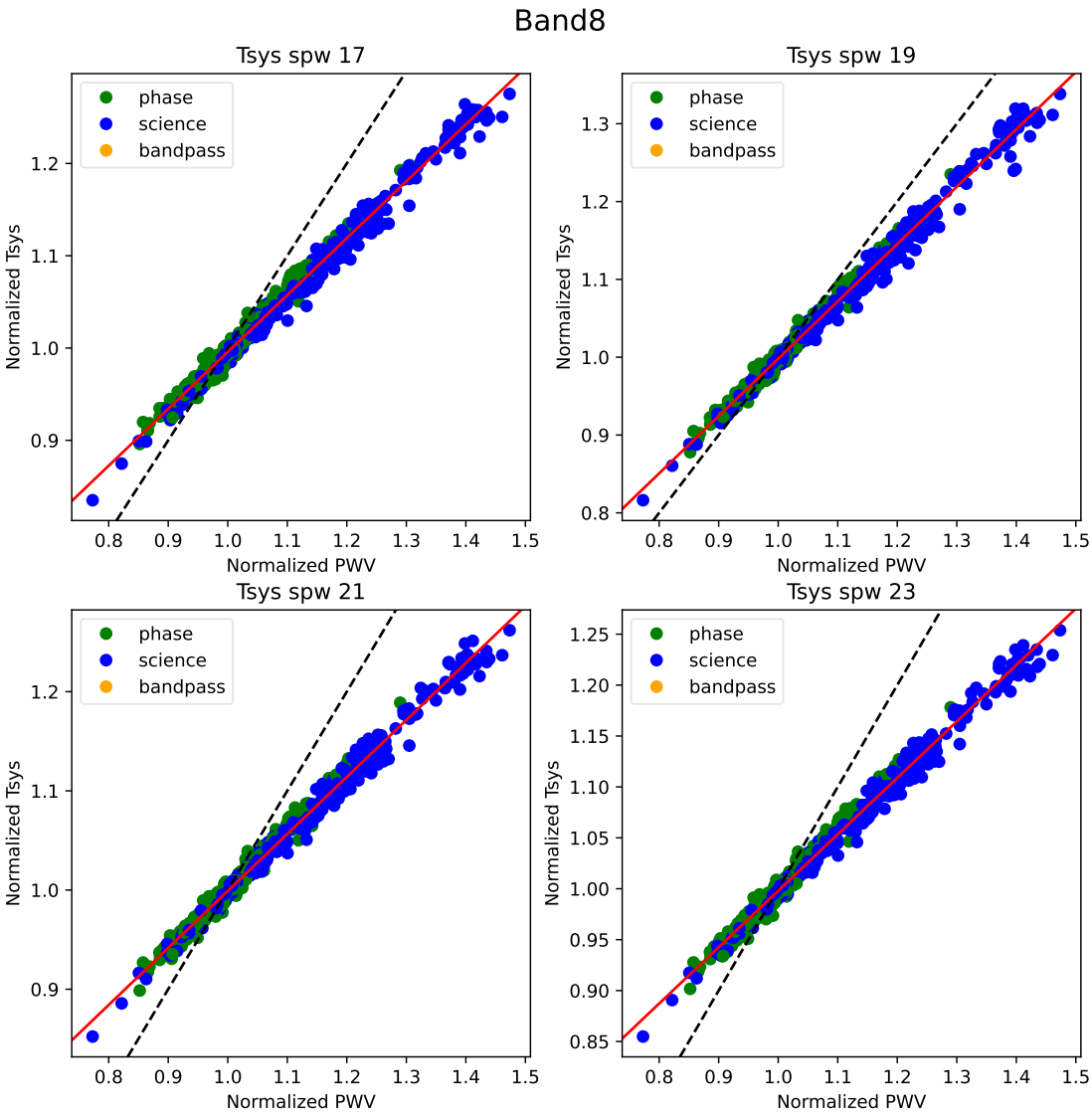

Figure 6. Correlation between the normalized $T_{\rm sys}$ and PWVs for different $T_{\rm sys}$ spws.

2.5. *Normalize Only to the Science target*

For some ALMA data, $T_{\rm sys}$ for the phase-cal target is not measured. Instead, the calibration uses the nearest science $T_{\rm sys}$ values as the phase-cal $T_{\rm sys}$. If we can normalize all $T_{\rm sys}$ values to the first $T_{\rm sys}$ of the science target instead of the first $T_{\rm sys}$ of each type of observing target itself, we can further reduce the number of $T_{\rm sys}$ measurements and thus no longer need to measure $T_{\rm sys}$ for phase-cal with our new method.

In this case, the normalized $T_{\rm sys}$ is calculated as

$$\begin{aligned} \hat{T}_{\rm sys,obs}(t) &= \frac{T_{\rm sys,obs}(t)}{T_{\rm sys,sci}(1^{\rm st})} \\ \hat{T}_{\rm WVR,obs}(t) &= \frac{T_{\rm WVR,obs}(t)}{T_{\rm WVR,sci}(1^{\rm st})} \end{aligned} \tag{15}$$

where $\hat{T}_{\rm sys,obs}(t)$ is the normalized $T_{\rm sys}$ averaged along the spectral axis and $\hat{T}_{\rm WVR,obs}(t)$ is the normalized $T_{\rm WVR}$. An example of $\hat{T}_{\rm sys}$ versus $\hat{T}_{\rm WVR}$ using the new normalization method is shown in Fig. 9. As we can see, all the data points follow almost the same linear correlation as the data normalized to each type of observing target. The bandpass data

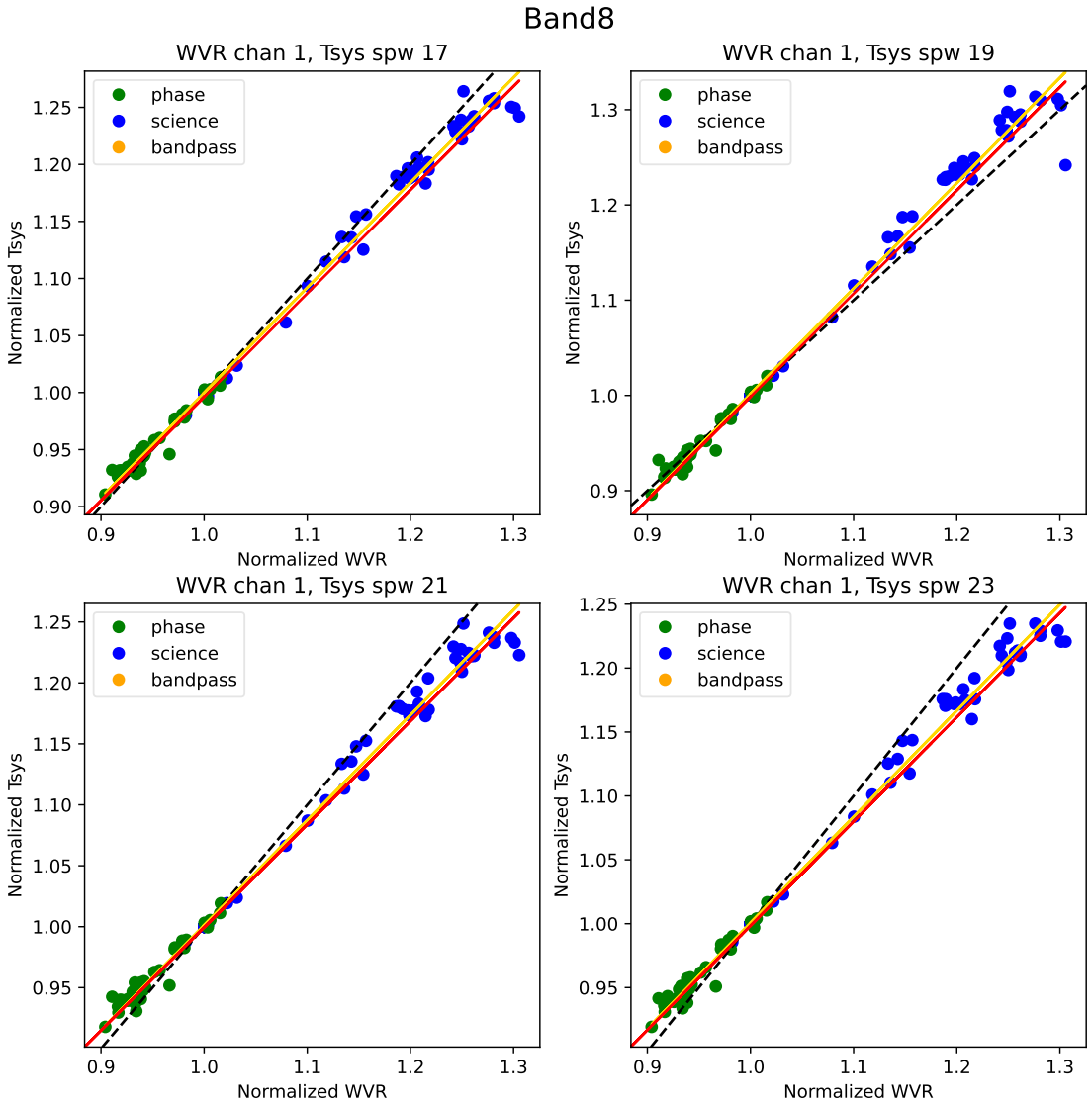

Figure 7. The correlation between the normalized $T_{\rm sys}$ and $T_{\rm WVR}$ for the 4 ATM scans we selected to fit the linear relation. The red and gold lines are the fitting relation with all ATM scans or just 4 ATM scans. The black dashed line indicates the 1-to-1 relation. We can see the fitting using all ATM scans are almost the same as just using 4 scans.

points follow the general trend but have a significant offset from other data points for spw 0. This is probably due to the different elevations of the bandpass and phase/science targets (see detailed discussion in Section 2.7. All the other measurement sets we have show that the normalized bandpass data lie along the same linear relation as the phase and science data (see another example in Fig. 25). We also test this relation in an extreme case for dataset Band10 where the science and phase-cal $T_{\rm sys}$ are drastically different, as shown in Fig. 10. We can see that even though the phase-cal target and science target are far apart in the sky, the two sets of data points still follow a quite similar tight linear relation. These tests show that we can further reduce the phase-cal and bandpass $T_{\rm sys}$ measurements as we can derive it from $T_{\rm sys}$ measurements for science target.

Note that for early ALMA cycles, the $T_{\rm sys}$ measurements are purely done for the phase-cal target. The $T_{\rm sys}$ for the science target is then assumed to be the same as the $T_{\rm sys}$ for the closest phase-cal scan. As we can see from this section, even though the phase-cal and science targets have different elevations, they generally follow the same $T_{\rm sys}$ vs $T_{\rm WVR}$ linear relation. Therefore, we can better extrapolate the science $T_{\rm sys}$ from the phase-cal $T_{\rm sys}$ using on the fitted linear relation.

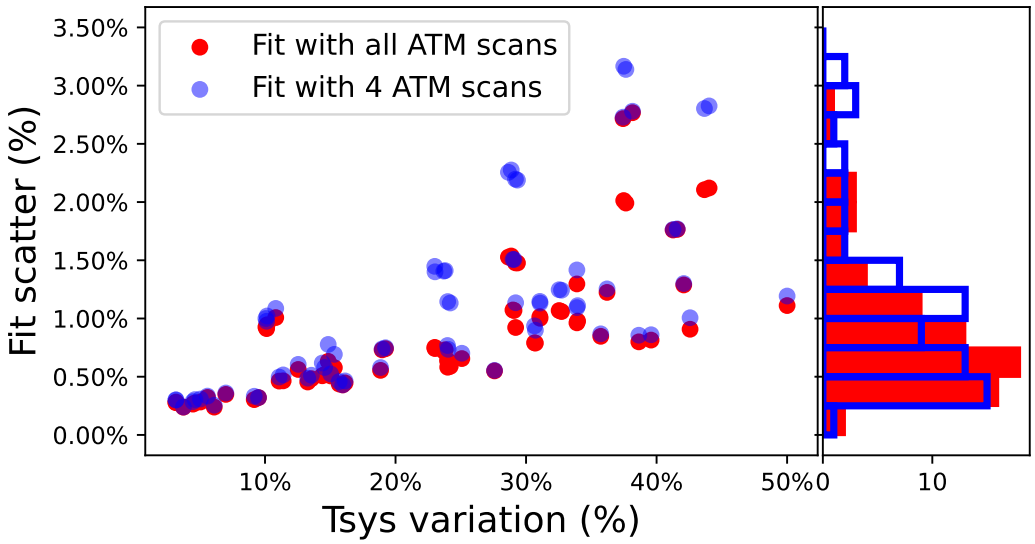

Figure 8. The scatter of data points around the $T_{\rm sys}$ vs $T_{\rm WVR}$ fits with all the antennas versus the maximal difference in $\hat{T}_{\rm sys}$ values for each $T_{\rm sys}$ spw of each dataset. The red and blue points are from fitting with all ATM scans or just 4 ATM scans respectively. The histogram at right side shows the distribution of the fit scatters using the two different methods. We can see the scatter of the fitting only increases slightly using just 4 ATM scans.

2.6. *Band 9 Data with Large $T_{\rm sys}$ Variation*

We would also like to test if this linear correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ holds in Band 9 and 10 data as they have a different receiver configuration (double-sideband) and generally higher atmospheric opacities. Dataset Band9b is ideal for this purpose as the $T_{\rm sys}$ variation is the largest among all of our datasets (50%). In Fig. 11, we show $T_{\rm sys}$ fitting and extrapolation using all ATM scans or just 4 ATM scans for one measurement set in this project. We can clearly see there is a difference in the fitting functions derived from all ATM scans or just 4 ATM scans. It seems the slope becomes steeper due to data points with higher $T_{\rm sys}$ values, which are not included if we use just 4 scans. This is also reflected in the extrapolation plot at right side of Fig. 11, as the predicted $T_{\rm sys}$ is lower than the measured $T_{\rm sys}$ for higher $T_{\rm sys}$ values.

Based on the analytical expression of $T_{\rm sys}$, we would expect a correlation that is actually curving up between $T_{\rm sys}$ and $T_{\rm WVR}$. Based on Eq. 2, 3 and 10, the correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ at low opacity limit can be simplified as

$$T_{\rm sys} \approx T_{\rm rx} + T_{\rm sky} = T_{\rm rx} + C_1 T_{\rm WVR} \tag{16}$$

where C is a constant. In this case, we expect linear correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ with positive intercept. However, at the high opacity limit, Eq. 3 is no longer a linear relation. Instead, the increase in $T_{\rm sys}$ is dominated by the increase in $\tau_{\rm sky}$ in the exponential form. Based on Eq. 2 and 3, we would expect that

$$T_{\rm sys} = \frac{T_{\rm sky}}{T_{\rm amb} - T_{\rm sky}} (T_{\rm sky} + T_{\rm rx}) \tag{17}$$

When $T_{\rm sky}$ is large, the increase in $T_{\rm sys}$ is dominated by $T_{\rm sky}^2/(T_{\rm amb} - T_{\rm sky})$ term, which has increasing slopes as $T_{\rm sys}$ increases. If we assume $T_{\rm sky}$ is still relatively proportional to $T_{\rm WVR}$, then we would find the correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ is turning more and more upward. In this case, the linear correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ we observed in other datasets should actually be a tangential line of the curving function. Because $T_{\rm sys}$ variations in other data sets are small, we can approximate the curving relation as a straight line. In Section 2.7, we also see a curving relation between $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$.

2.7. *Atmospheric modeling for $T_{\rm sys}$ versus $T_{\rm WVR}$*

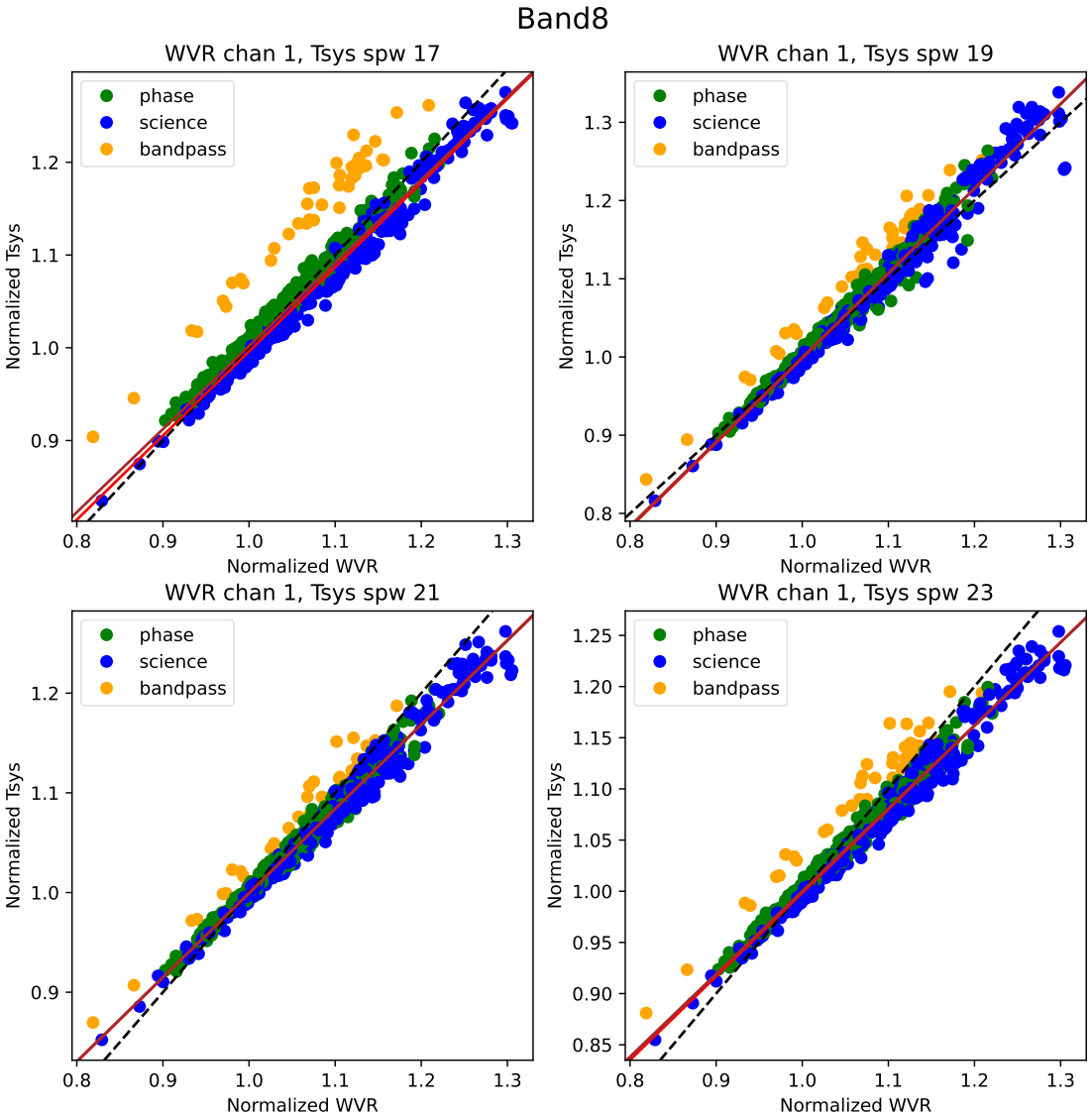

Figure 9. $T_{\rm sys}$ versus $T_{\rm WVR}$ normalized to the first science ATM scan for the dataset Band8. The black dashed line is the 1-to-1 relation. The brown solid line is the linear fitting to the data excluding the bandpass data. The red solid line is the original fitting relation to the data normalized to each type of observing target. We can see the two fitting relation are almost the same. The bandpass data for spw 0 has a significant offset from the fitted relation, which we think is probably due to the incorrect measurement of $T_{\rm sys}$ at that time. See Fig. 25 in Appendix A for a good example.

In the previous sections, we fit $T_{\rm sys}$ vs $T_{\rm WVR}$ heuristically and uses the fitted relation to extrapolate the $T_{\rm sys}$ continuously. We find in most cases the correlation between $T_{\rm sys}$ and $T_{\rm WVR}$ is linear. However, the exact slopes and intercepts of the correlations vary across different frequencies. Although we can use less than 4 ATM scans to fit the relation for each spw in each dataset, there might be cases when our selected ATM scans have similar $T_{\rm sys}$ values, and hence might give us inaccurate fitting relation among small $T_{\rm sys}$ ranges. Furthermore, the linear relation breaks up when $T_{\rm sys}$ variations become really large (e.g. dataset Band9b). A more robust method is to use atmospheric transmission at microwaves (ATM) modeling (Pardo et al. 2001) to predict the $T_{\rm sys}$ versus $T_{\rm WVR}$ relation in various cases. In this section, we try the ATM modeling for dataset Band8 to see if the ATM modeling can give us as good a fit as our heuristic method.

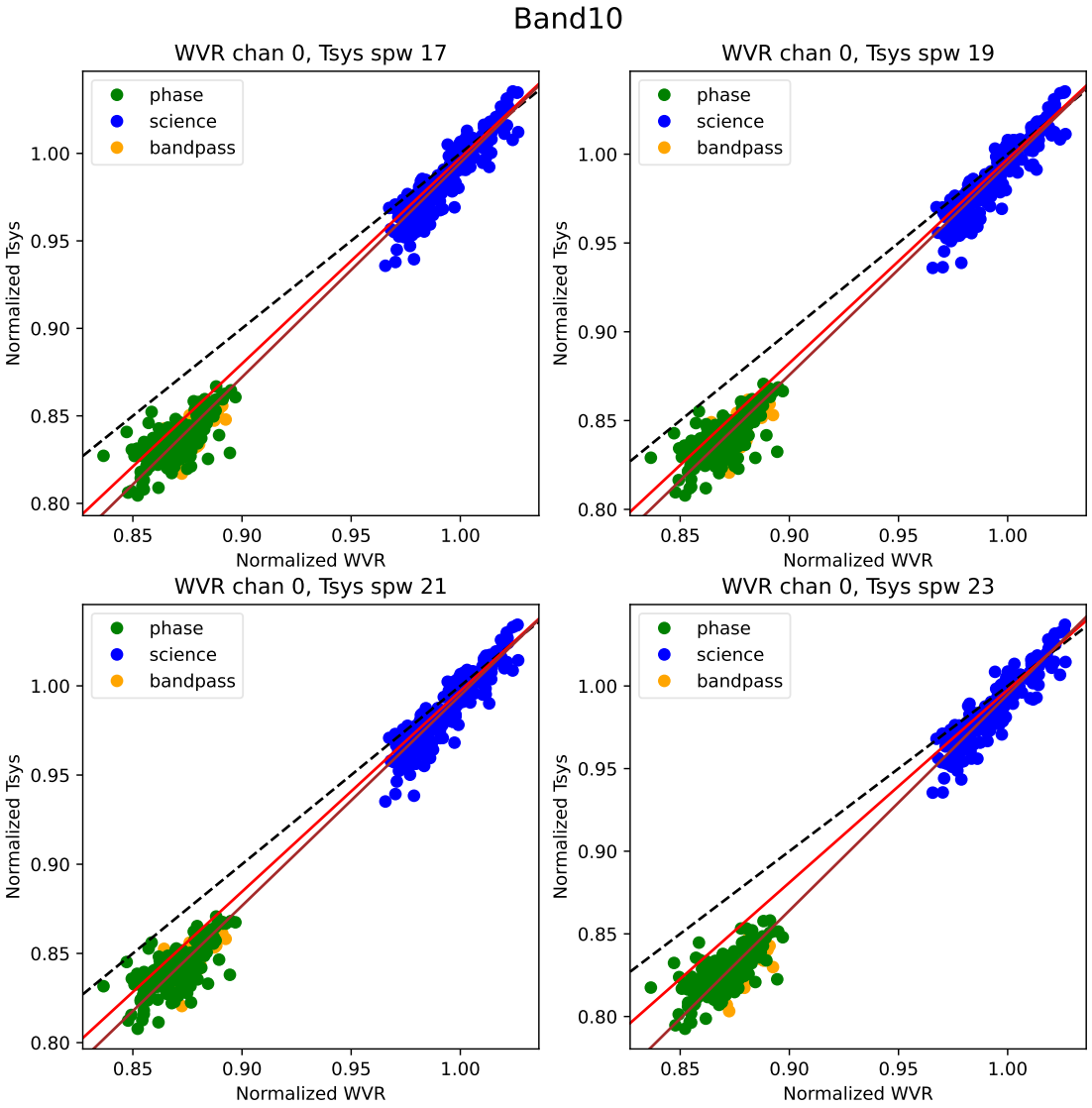

Figure 10. $T_{\rm sys}$ versus $T_{\rm WVR}$ normalized to the first science ATM scan for the Band 10 data. The black dashed line is the 1-to-1 relation. The maroon solid line is the linear fitting to the data. The red line is the original fitting relation to the data normalized to each type of observing target.

We use the CASA function `aU.plotAtmosphere` to generate $T_{\rm sys}$ and $T_{\rm WVR}$ spectra at given frequency ranges. We set most of the parameters to the default for the ALMA site (height 5000 m, pressure 557 mb and temperature 274 K). For dataset Band8, the elevation of the science target is about 53 degrees and the median PWV value is 0.85 mm (Table 2). We hence input different elevations and PWVs centered around these values to generate the $T_{\rm sys}$ and $T_{\rm WVR}$ spectra. Note that `aU.plotAtmosphere` only outputs $T_{\rm sky}$ and zenith opacity τ_0 ($\tau_{\rm sky} = \tau_0 \sec z$) where z is the elevation). Therefore, we use Eq. 1 to convert these values to $T_{\rm sys}$ for each given $T_{\rm rx}$ and elevation angle. An example $T_{\rm sys}$ spectrum is shown in Fig. 12 assuming $T_{\rm rx}$ of 100 K and elevation angle of 53 degrees. As we can see, the shape of $T_{\rm sys}$ spectrum does not vary much for different PWVs and elevations. However, it is clear that varying PWVs or elevations can have slightly different effects on the shape of the spectrum. If we increase PWV values, broader part of the spectrum (due to H_2O) increases while the narrow components due to Ozone lines do not. On the other hand, if we increase the elevation, it will increase the strengths of both continuum and Ozone lines. However, the Ozone lines are narrow and have a relatively small effect on the $T_{\rm sys}$ values.

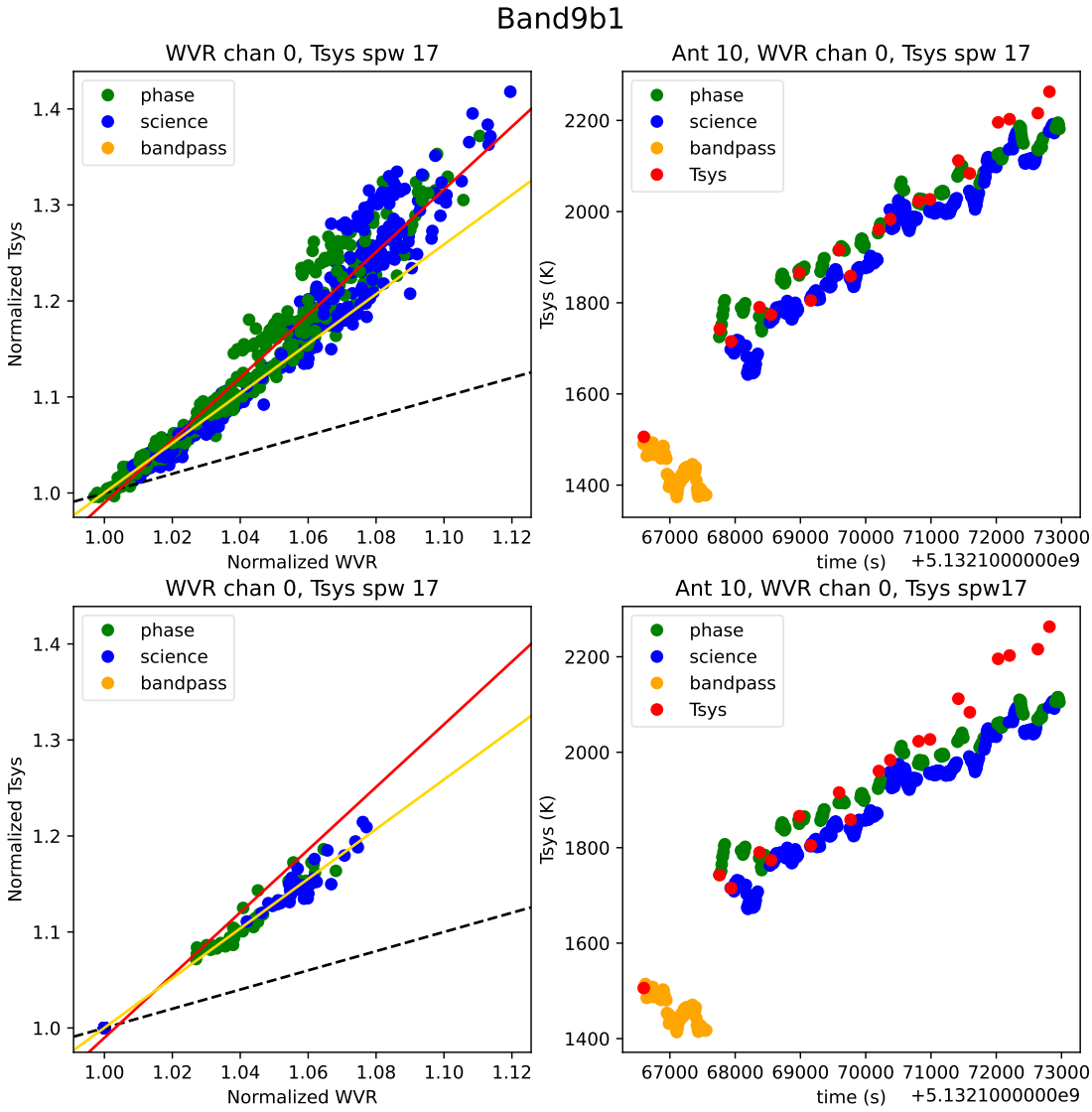

Figure 11. (Left) $T_{\rm sys}$ versus $T_{\rm WVR}$ (channel 0) for one data set in Band 9 project 2019.1.00013.S. The upper left panel shows $\hat{T}_{\rm sys}$ versus $\hat{T}_{\rm WVR}$ in spw 17 using all ATM scans while the lower left panel shows the correlation with selected 4 ATM scans. The red and yellow lines are the fits derived using all ATM scans and just 4 ATM scans. (Right) The measured and extrapolated $T_{\rm sys}$ for different observing targets as a function of time. Red points are measured $T_{\rm sys}$ values. $T_{\rm sys}$ in upper panel is extrapolated based on fits with all ATM scans while $T_{\rm sys}$ in lower panel is extrapolated based on fits using only selected 4 ATM scans.

To explore the $T_{\rm sys}$ versus $T_{\rm WVR}$ correlation, we average the $T_{\rm sys}$ and $T_{\rm WVR}$ across each spw. We first explore how $T_{\rm sys}$ is correlated with $T_{\rm WVR}$ if we just change PWV values. Fig. 13 shows our modeling overlaid by the acutual normalized data for $T_{\rm sys}$ spw 17 and WVR channel 1. For the modeling, we vary the $T_{\rm rx}$ in the left panel and elevations in the right panels to see how different parameter choices affect our modeling slopes. For dataset Band8, $T_{\rm rx}$ varies from 50 to 130 K for different antennas while elevations of the science and phase targets are around 50 – 55 degrees. As we can see, the ATM modeling shows a slight non-linear curvature for a high range of $T_{\rm sys}$, which depends slightly on the assumed $T_{\rm rx}$ and the elevation. However, the modeling curve is generally within the range of the data scatter. Therefore, we can still fit a linear relation and extrapolate $T_{\rm sys}$ with a good accuracy.

We then explore how elevations could affect the $\hat{T}_{\rm sys}$ vs $\hat{T}_{\rm WVR}$ correlation. In Fig. 14, we can see the slopes of $\hat{T}_{\rm sys}$ vs $\hat{T}_{\rm WVR}$ are much steeper than is seen in actual data. This is what we expect since reducing elevations or increasing the airmass will increase both the wet and dry components in the same way. While $T_{\rm sys}$ has comparable contributions from

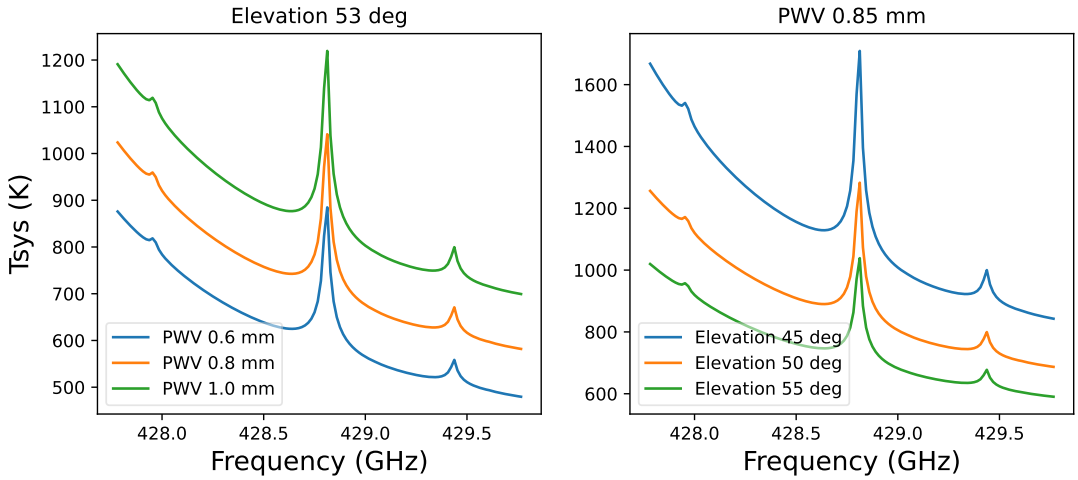

Figure 12. The ATM modeling $T_{\rm sys}$ spectrum for different PWV (left) and elevation (right) values for dataset Band8 spw 17.

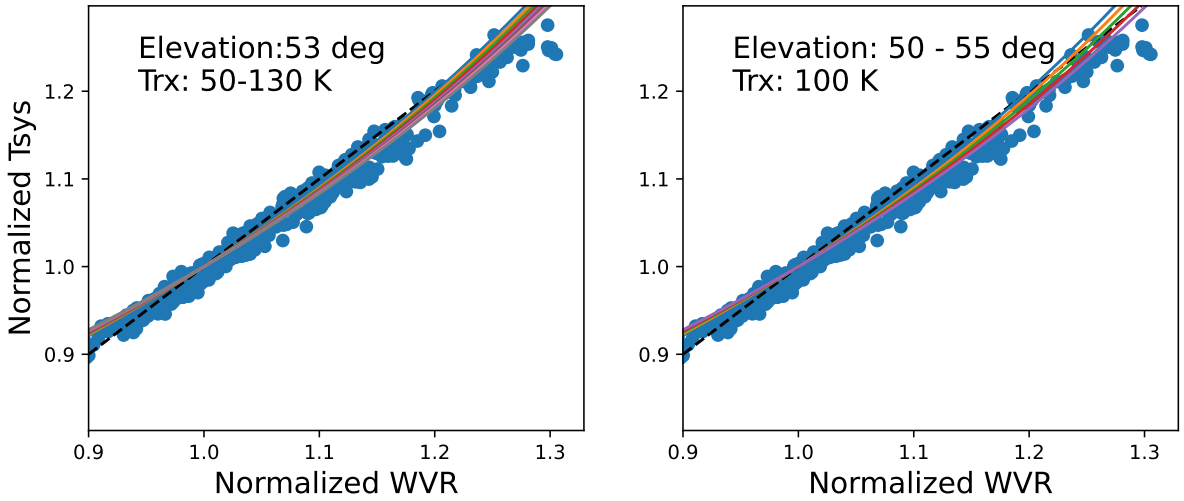

Figure 13. The correlation between $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ (Fig. 3) overlaid by the predicted correlation curves from ATM modeling. $T_{\rm sys}$ is from spw 17 and $T_{\rm WVR}$ is from channel 1. The left panel shows curves for different $T_{\rm rx}$ values, while the right panel has curves for different (but fixed) elevations of the target.

wet and dry components, $T_{\rm WVR}$ is dominated by the wet component. Therefore, the increase in $T_{\rm sys}$ will be steeper than that of $T_{\rm WVR}$. In our actual observations, the elevation change is quite small, generally 1 – 2 degrees, hence the change in $T_{\rm sys}$ and $T_{\rm WVR}$ should be dominated by PWV variations. For dataset Band8, we also see the difference between normalized bandpass data and science/phase data in Fig. 9. This difference could also be explained by the elevation difference between the bandpass target and phase/science target. In Fig. 14, the thick red and purple lines show the ATM modeling of two fixed elevations of 53 and 48 degrees with varying PWVs. We can see the difference in elevations could well explain the offsets between the normalized bandpass and phase/science data.

3. AC & SQLD DATA IN TRACKING $T_{\rm sys}$

As mentioned in Section 1.4, we can also explore whether to use AC or SQLD data to track $T_{\rm sys}$ variation. Since AC and SQLD data are equivalent to one another (see Fig. 1 right panel), we only need to compare $T_{\rm sys}$ with one of the two quantities. We use AC data for comparison since the data size is much smaller. We use a similar method to normalize the AC data and compare it with the normalized $T_{\rm sys}$. Fig. 15 shows the comparison between $T_{\rm sys}$ versus

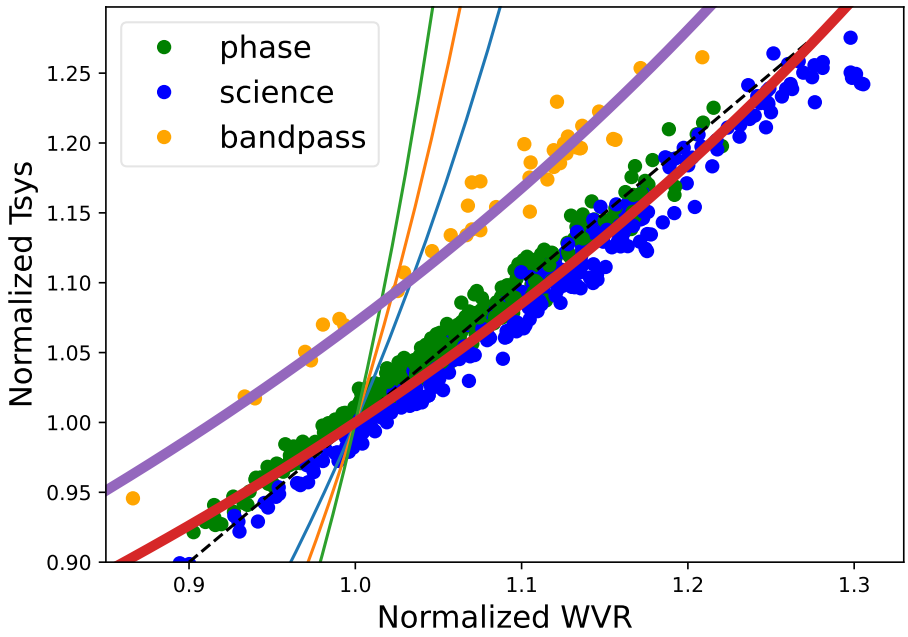

Figure 14. The correlation between $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$ (Fig. 9) overlaid by ATM modeling curves (blue, orange and green) obtained by varying the elevation (from 40-60 degrees) at a constant PWV.. Note in this case data from dataset Band8 is normalized only to the science target to show how elevation affects the data. The 3 ATM modeling curves correspond to different PWV values (blue: 0.6 mm; orange: 0.8 mm; green: 1.0 mm). The thick red line is the ATM modeling by changing PWV values from 0.5 to 1.0 mm with a fixed elevation of 53 degrees. The purple thick line is also from ATM modeling by changing PWVs with a fixed elevation of 48 degrees. Note that the purple line is normalized to the same value as the red line instead of the value from its own target for ease of comparison between absolute values in these two cases. $T_{\rm rx}$ in the modeling is set to be 100 K.

AC and $T_{\rm sys}$ versus WVR correlation for dataset Band8 and Band7a. We can see that the AC data also has a tight correlation with $T_{\rm sys}$ for dataset Band7. However, the correlation does not work well for dataset Band8 with a large scatter. Therefore, we cannot just fit the relation to several ATM scans to calculate the continuous $T_{\rm sys}$ with good precision.

The other thing we find is that AC and $T_{\rm sys}$ do not follow the proportional correlation as might be expected. The major reason is that AC data track the total signal received after the atmosphere attenuation while $T_{\rm sys}$ tracks the total signal before it comes through the atmosphere, as shown in the diagram in Fig. 17. The AC data is directly proportional to the total signal received by the antenna, which is mainly comprised of emission from the sky ($\eta_{\rm eff}T_{\rm sky}$), the receiver itself ($T_{\rm rx}$), and other fixed losses terminating at ground ($(1-\eta_{\rm eff})\times T_{\rm amb}$). However, based on Eq. 1, $T_{\rm sys}$ is not directly proportional to these 3 components added together. Instead, $T_{\rm sys}$ can be thought of as brightness temperature of a fake source in space that generates a signal equal to the 3 components added together after atmosphere attenuation. In other words, for a single band setting,

$$P_{\rm AC} \propto \eta_{\rm eff} e^{-\tau_0 \sec z} T_{\rm sys} = T_{\rm rx} + \eta_{\rm eff} T_{\rm sky} + (1-\eta_{\rm eff}) \times T_{\rm amb} \tag{18}$$

We call the right side of the equation attenuated $T_{\rm sys}$. We also normalize the attenuated $T_{\rm sys}$ the same way as we do for the original $T_{\rm sys}$ and compare it with normalized AC data. The comparison between $T_{\rm sys}$ and attenuated $T_{\rm sys}$ versus AC correlation is shown in Fig. 16. As we can see, the normalized attenuated $T_{\rm sys}$ follows the 1-to-1 relation with normalized AC as suggested by Eq. 18. If we rearrange Eq. 18, it becomes

$$T_{\rm sys} \propto \frac{P_{\rm AC}}{e^{-\tau_{\rm sky}}} = \frac{P_{\rm AC}}{1 - T_{\rm sky}/T_{\rm amb}} \tag{19}$$

Therefore, even though we have the AC data, we still need a method to continuously monitor $T_{\rm sky}$ to get $T_{\rm sys}$. Additionally, the AC and SQLD data suffer from gain drift and gain step changes between scans, as noted above and seen in Fig. 1. Thus we cannot measure continuous $T_{\rm sys}$ just with the AC data.

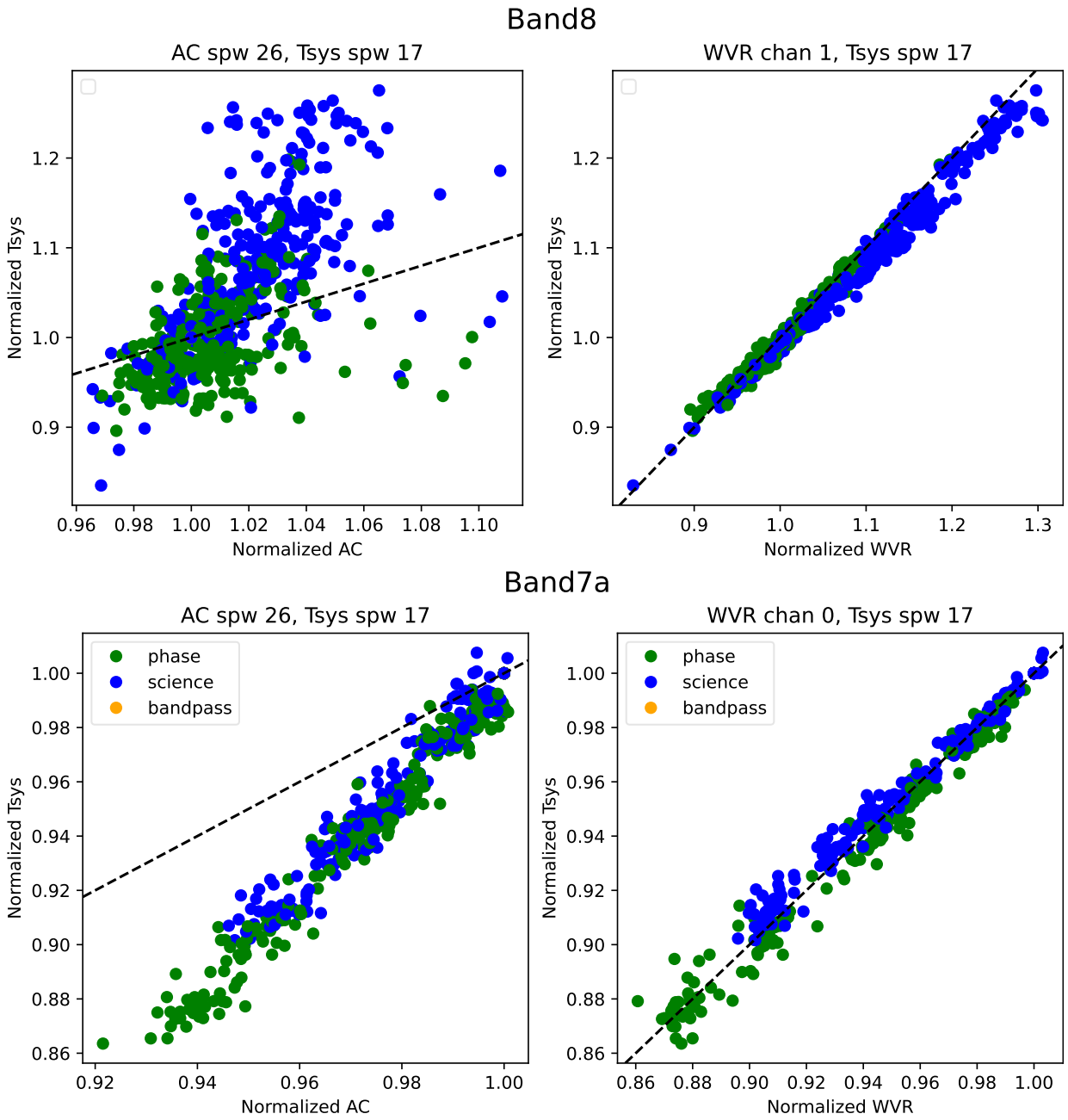

Figure 15. The correlation between the $\hat{T}_{\rm sys}$ and matched $\hat{T}_{\rm WVR}$ and normalized $A_{\rm AC}$ for all the antennas. Both WVR and auto-correlation data is averaged over 10 seconds.

4. APPLYING CONTINUOUS $T_{\rm sys}$ TO THE CALIBRATION

In Section 2, we demonstrated the viability to use WVR data to track $T_{\rm sys}$ continuously. In this section, we apply the extrapolated continuous $T_{\rm sys}$ in calibration to test whether our new method for measuring $T_{\rm sys}$ works. We calibrate each dataset with the original $T_{\rm sys}$ table, the new continuous $T_{\rm sys}$ table extrapolated using all ATM scans and that using just 4 ATM scans with CASA (McMullin et al. 2007) package. We then make images from data calibrated using these 3 different methods and see if the measured fluxes for the same target are more consistent with each other using our new methods. The detailed description of the scripts we use for the data processing can be found at https://github.com/heh15/ALMA_intern_Tsys.git.

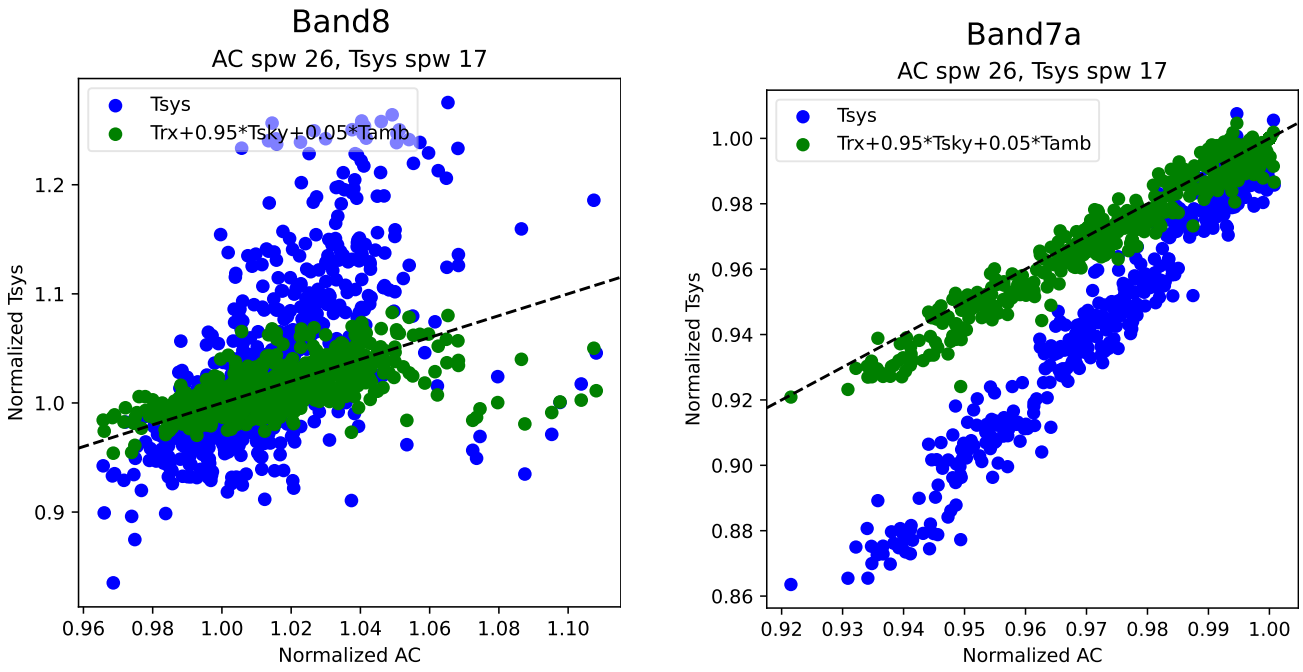

Figure 16. The correlation between the normalized $T_{\rm sys}$ (blue) and attenuated $T_{\rm sys}$ (green) and normalized AC data for the Band 8 and 7 data we have for all antennas of one spectral window. As we can see, the attenuated $T_{\rm sys}$ follows the 1-to-1 correlation with AC data, which proves equation 18 to be right.

4.1. *Creation of $T_{\rm sys}$ Table*

In this subsection we discuss how we construct the new $T_{\rm sys}$ table used for the calibration. We note that the original $T_{\rm sys}$ table used for calibration is a spectrum with two polarizations. Recording all the extrapolated $T_{\rm sys}$ spectra in one big table would take a lot of disk space. Based on our check of the $T_{\rm sys}$ spectrum plots generated using the original calibration script, the shape of the $T_{\rm sys}$ spectrum of each spectral window does not vary much as a function of time. Therefore, we can just record the initial $T_{\rm sys}$ for each observing target in the $T_{\rm sys}$ table and record the ratio of the extrapolated $T_{\rm sys}$ relative to the initial $T_{\rm sys}$ into an amplitude gain table. In this case, the two tables we provide for $T_{\rm sys}$ calibration are

$$\begin{aligned} T_{\rm sys}(t,\nu) &= T_{\rm sys,obs}(1^{\rm st},\nu) \\ G(t) &= \sqrt{1\Big/\left[\frac{T_{\rm sys}(t)}{T_{\rm sys}(1^{\rm st})}\right]_{\rm fit}} \end{aligned} \tag{20}$$

where $T_{\rm sys}(t,\nu)$ is the recorded $T_{\rm sys}$ spectrum as a function of time t and frequency ν, and G is the derived gain as a function of t, $T_{\rm sys,obs}(1^{\rm st},\nu)$ is the first $T_{\rm sys}$ spectrum measured for each type of observation of given antenna and spectral window and $\left[\frac{T_{\rm sys}}{T_{\rm sys}(1^{\rm st})}\right]_{\rm fit}$ is the extrapolated normalized $T_{\rm sys}$ from the fitting. We note that G is not directly equal to the $\hat{T}_{\rm sys}$ values. This is due to the different methods that CASA uses to handle $T_{\rm sys}$ and gain table. For each baseline, the correlated amplitude is

$$S(i,j) \propto \sqrt{T_{\rm sys}(i)T_{\rm sys}(j)} \propto \frac{1}{G(i)G(j)} \tag{21}$$

Therefore, the G is written so that it can be properly translated to the variation in $T_{\rm sys}$.

In Section 2.4, we tested using only 4 $T_{\rm sys}$ measurements to fit the linear relation between $\hat{T}_{\rm sys}$ and $\hat{T}_{\rm WVR}$. We saw that the difference between this method and using all $T_{\rm sys}$ measurements is small. However, we still need to quantify if the small difference in the linear fits makes much difference in the measured flux of the image product. In this case, we

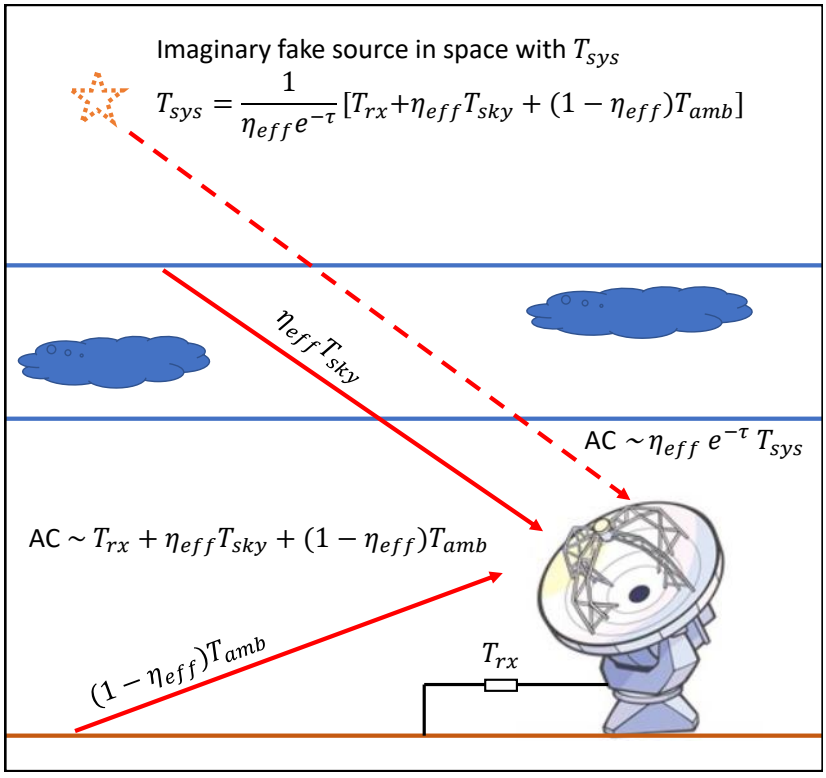

Figure 17. Illustration of the theoretical relationship between the autocorrelation signal AC and $T_{\rm sys}$. AC is proportional to the total power signal received by an antenna, which includes receiver noise ($T_{\rm rx}$), sky noise ($T_{\rm sky}$), and thermal noise due to losses and spillover terminating around the ambient temperature ($(1 - \eta_{\rm eff}) \times T_{\rm amb}$). However, $T_{\rm sys}$ is corrected for a source outside the atmosphere, and includes an extra term due to the atmospheric attenuation.

also apply Eq. 20 to create the alternative $T_{\rm sys}$ table with the fitting relation derived from 4 ATM scans. Therefore, we also extrapolate $T_{\rm sys}$ based on the fitting relation from 4 ATM scans.

4.2. *Calibrating and Imaging the Data*

After we create the $T_{\rm sys}$ table, we then apply the continuous $T_{\rm sys}$ in calibration. The calibration script we use is generated from the command `es.generateReducScript()`. We then modify the script to use the alternative $T_{\rm sys}$ and amplitude gain tables created (see details in https://github.com/heh15/ALMA_intern_Tsys.git). We also run the original calibration script to calibrate the data with the original discrete $T_{\rm sys}$ method for comparison.

After calibration, we then proceed with making continuum images. We generally adopt the default settings using the command `tclean`. We set the `robust` parameter to be 2.0 instead of the default 0.5 to maximize the sensitivity and hence flux accuracy. We also set the number of iterations to be 0 to only make the dirty image. This process will reduce any effects from `tclean` itself when making comparisons between fluxes from different calibration datasets. For projects with multiple datasets, we directly compare the measured fluxes from different datasets to see if they are consistent with each other. For projects with just one measurement set, we further make images using just the first half or the second half of the science scans and then compare fluxes among these 3 images. The top panel in Fig. 18 shows an example of images made with the 3 different methods for dataset Band8. We can see that the structure of these images looks almost the same. This is what we expect since $T_{\rm sys}$ should only affect the intensity scale of the final image. The bottom panel of Fig. 18 shows the images of the same target made with all scans or just half of the scans using the new method to extrapolate $T_{\rm sys}$ from all ATM scans. We can see the structures of these images are also almost the same, which is also what we expect as images from the same dataset should have the similar uv coverage. On the other hand, if we compare images made from different datasets, we can see there is a larger difference in image structure. This is most obvious in the Band 9 project 2019.1.00013.S shown in Fig. 19. The beam sizes for the 3

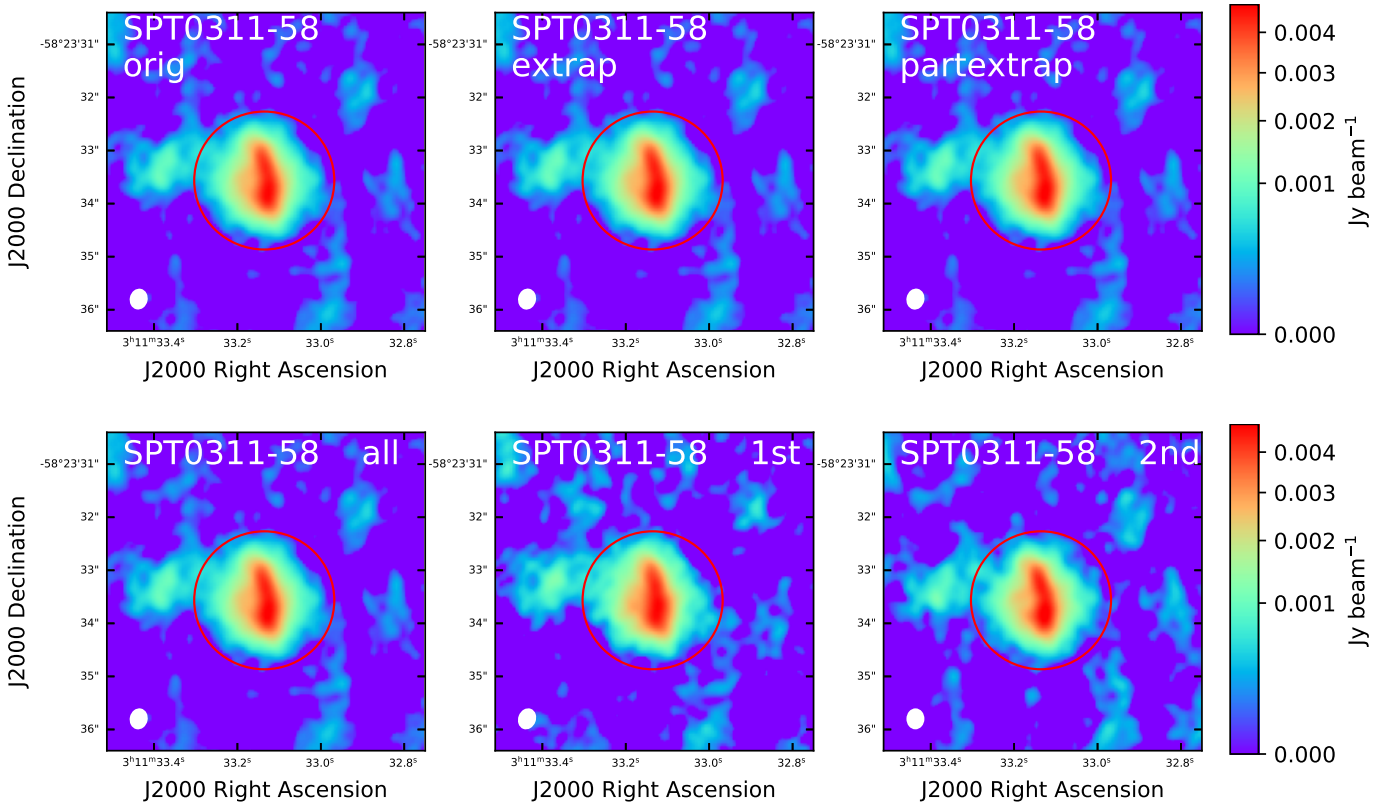

Figure 18. (Top) dirty images made using originally measured $T_{\rm sys}$ (orig), extrapolated continuous $T_{\rm sys}$ with all ATM scans (extrap) and that with just 4 ATM scans (partextrap) for dataset Band8 of SPT0311-58. The red circle is the aperture used for flux measurements. The fluxes for these 3 images are 0.0425, 0.0429 and 0.0407 Jy. (Bottom) dirty images made using alternative continuous $T_{\rm sys}$ table derived from fitting all ATM scans. The 3 columns are images made using all, 1st half and 2nd half of the science scans. The fluxes for these 3 images are 0.0429, 0.0432 and 0.0428 Jy.

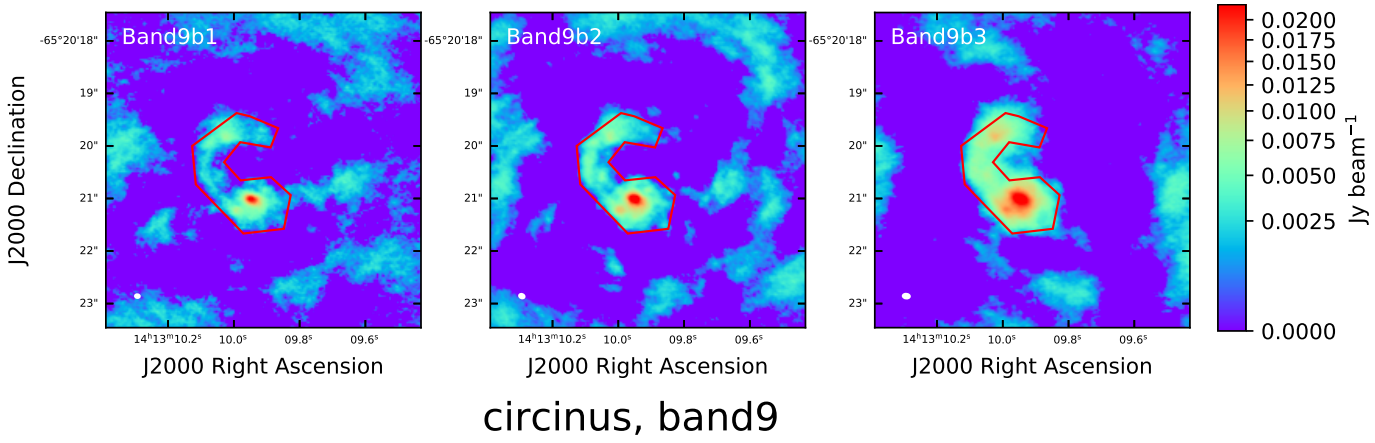

Figure 19. Dirty images made using alternative continuous $T_{\rm sys}$ table using all ATM scans for the 3 data sets in the Band 9 project 2019.1.00013.S. The red polygons are apertures used to measure the flux. The fluxes for these 3 images are 0.875, 1.0 and 1.35 Jy.

different images are also significantly different from each other. For dataset Band7b, all images have similar beam shapes so differences among the image structures are not that significant.

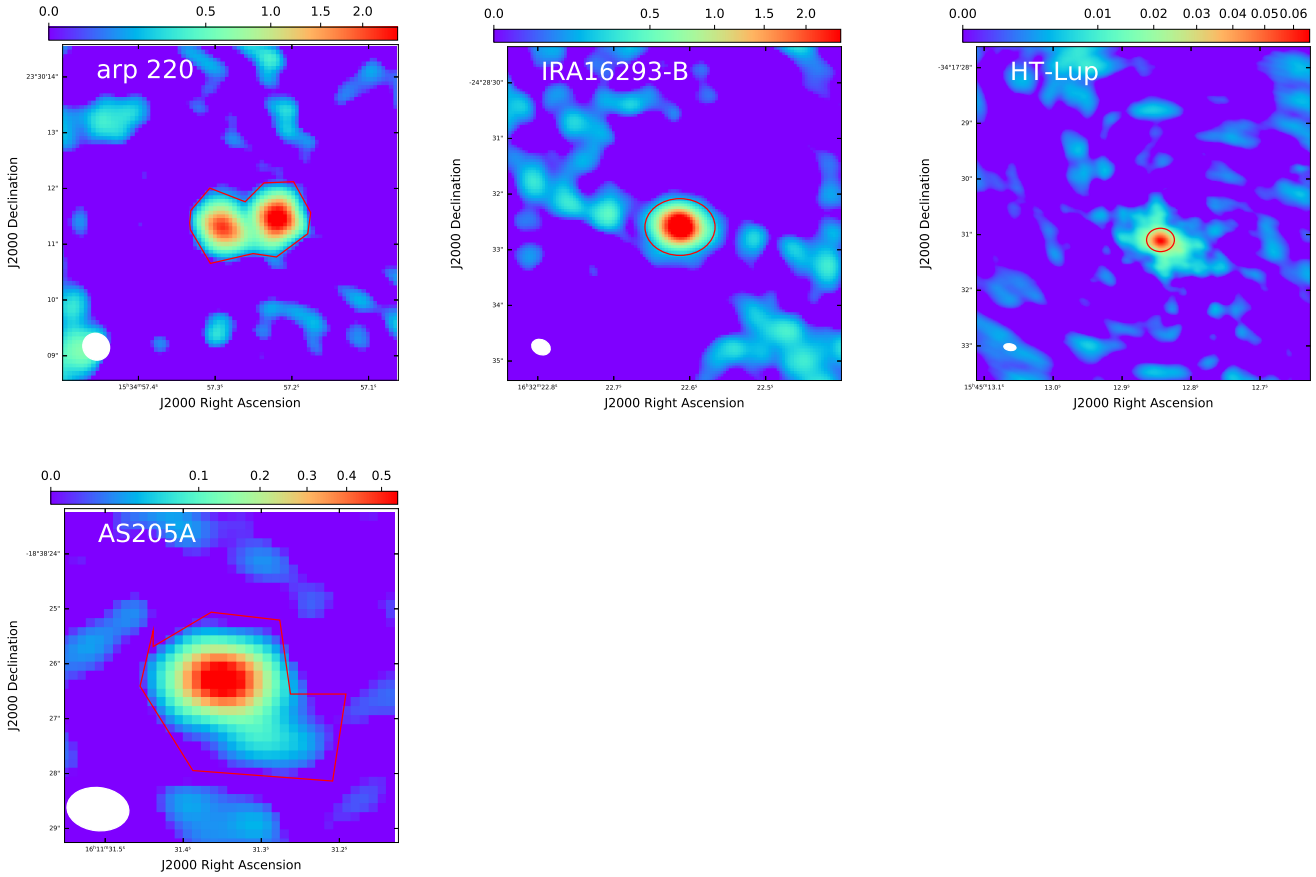

Figure 20. Dirty images for other sources made using alternative continuous $T_{\rm sys}$ table using all ATM scans. For Band 7 project 2018.1.01210.S (AS205A), we show the image made from data uid://A002/Xda1250/X2387 (Band7b1). The red polygons are apertures used to measure the flux.

Once the images are made, we draw an aperture around the central point source to measure the flux. Since the change in $T_{\rm sys}$ does not change the structure of the continuum image, we can compare the fluxes measured from the same aperture for the same target (as long as the aperture is not missing any flux, or there is no decorrelation – as decorrelation will reduce the total flux potentially). The apertures we used to measure the fluxes are shown in Fig. 18, 20 and 19. Note that as well as changes in $T_{\rm sys}$, changes in phase decorrelation during a single observation and between observations may also affect the measured fluxes, and account for some of the scatter in fluxes in Table 3, 4 and 5. However, we assume this effect is the same is all reductions, independent of the $T_{\rm sys}$ calibration method.

4.3. *Flux Comparison*

The measured fluxes are recorded in Table 3, 4 and 5. The flux uncertainty can be separated into two parts, the measurement error and the calibration error. The measurement error is calculated as

$$\text{Meas.Err} = \text{rms}\sqrt{N_{\rm beam}} \tag{22}$$

where the rms is the measured noise of the image and $N_{\rm beam}$ is the number of beams across the aperture used to measure the flux. On the other hand, our alternative method to measure $T_{\rm sys}$ should mainly work on reducing the calibration error. To test if our new method improves the flux calibration accuracy, we need to quantify the calibration error for each method we use. For projects with single dataset, we compare fluxes of images made with 1st half, 2nd

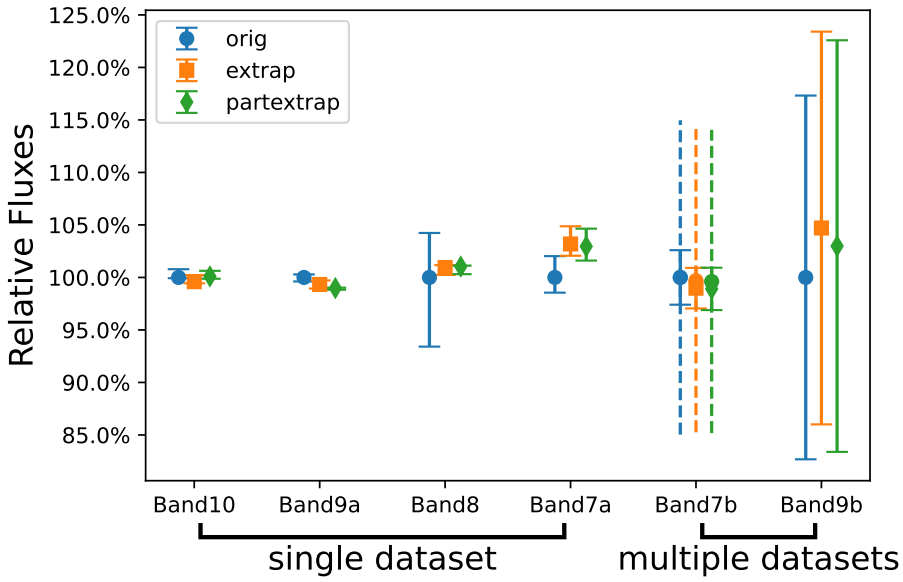

Figure 21. Comparison of fluxes of images made using the original $T_{\rm sys}$ (orig), continuous $T_{\rm sys}$ extrapolated from all ATM scans (extrap) and that extrapolated from 4 ATM scans (partextrap). The vertical axis are the ratio of fluxes to the fluxes using original method. The first 4 projects contain single dataset while the last two projects contain multiple datasets. The dashed line for Band7b project shows the error including the two datasets that has abnormal flux values (see Table .4). We can see the flux consistency is generally better for most of the projects with smaller uncertainties. See description in Section 4.3.

half and all scans and calculate the maximal difference between the 3 flux values as the calibration error. For projects with multiple datasets, the calibration error is calculated as the standard deviation of fluxes of the different datasets. Both measurement and calibration errors are recorded in Table 3, 4 and 5.

To further compare how our methods work for datasets from different frequency bands, we normalize the flux values to the flux value using the original discrete $T_{\rm sys}$ calibration method with all ATM scans. The relative uncertainties are calculated as the calibration error divided by the flux value for each method using all ATM scans. The comparison is shown in Fig. 21. We can see that fluxes using the original $T_{\rm sys}$ table do not differ significantly from the fluxes using alternative $T_{\rm sys}$ and gain tables, with maximal differences smaller than 5 %. We also demonstrate that the extrapolated $T_{\rm sys}$ using 4 ATM scans gives us fluxes that are almost the same as when using all ATM scans, which proves it is viable to significantly reduce the number of $T_{\rm sys}$ measurements by using WVR-tracked $T_{\rm sys}$. Furthermore, it seems for most of the datasets, our new methods give better flux consistency, especially for dataset Band8 which has the highest PWV value of 0.85 mm among all the datasets. As shown in Fig. 4, $T_{\rm sys}$ variation throughout the observation is dramatic without a straightforward pattern (monotonously increasing or decreasing). In this case, the discrete $T_{\rm sys}$ measurements probably did not catch the fluctuations of the real $T_{\rm sys}$. Our new method instead catches the variation in $T_{\rm sys}$ between the discrete ATM calibrations, and thus keeps the flux consistent. This method also works for dataset Band7b with multiple data sets with relative flux uncertainties reduce from 2.5% to 1.9%. The only project that gives us larger flux uncertainties using our new methods is the dataset Band9b. For this project, the uncertainties using all 3 methods are quite large ($\sim$ 15%). The large uncertainties are probably due to different uv-coverages, the complex target structure and relatively high phase noise in these datasets. This is also a tricky dataset for which the linear fitting does not work as well as for other data sets. In our future work, we will explore if the larger uncertainty is caused by imperfect fitting of $T_{\rm sys}$ versus $T_{\rm WVR}$ relation.

4.4. *Additional considerations for the continuous $T_{\rm sys}$ method*

For some targets, the source brightness temperature in single-dish measurements can be significant compared with $T_{\rm sys}$. For example, this may occur for bright galactic targets in 12CO, bright masers, or for some Solar System objects in continuum. As the widths of galactic spectral lines are generally negligible (¡1%) compared with the normal bandwidth (2GHz) used to measure Tsys, and the continuous $T_{\rm sys}$ method uses a spectrally-averaged broad-band Tsys,

Table 3. Flux measured for project with only one dataset

Data Label	Target	Scans used	Flux (Jy)			Meas. Err. (Jy)	Beam (")
			Tsys_orig	Tsys_extrap	Tsys_partextrap		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Band10	Arp 220	all	7.169	7.146	7.187	0.39	0.5 x 0.47
		1st half	7.1638	7.135	7.171	0.42	0.52 x 0.45
		2nd half	7.1077	7.09	7.133	0.4	0.48 x 0.48
		MAX. DIFF[a]	0.06	0.056	0.054		
Band9a	IRAS16293-B	all	10.43	10.36	10.324	0.34	0.34 x 0.27
		1st half	10.4	10.32	10.32	0.36	0.35 x 0.27
		2nd half	10.47	10.4	10.34	0.38	0.34 x 0.26
		MAX. DIFF.	0.07	0.04	0.02		
Band8	SPT0311-58	all	0.0425	0.0429	0.0429	0.0015	0.35 x 0.3
		1st half	0.0453	0.0432	0.0433	0.0019	0.36 x 0.29
		2nd half	0.0407	0.0428	0.0429	0.0017	0.35 x 0.3
		MAX. DIFF	0.0046	0.0003	0.0003		
Band7a	HT-Lup	all	0.173	0.178	0.178	0.0038	0.22 x 0.12
		1st half	0.169	0.175	0.175	0.0044	0.22 x 0.12
		2nd half	0.175	0.18	0.18	0.0045	0.22 x 0.12
		MAX. DIFF.	0.006	0.005	0.005		

Columns: (1) The label of each dataset (see Table 2) (2) The target name. (3) The science scans used to make images. (4) Fluxes of the source with images made using original calibration script. (5) Fluxes of the source with images made using modified script with alternative $T_{\rm sys}$ table. The $T_{\rm sys}$ is extrapolated based on all ATM scans. (6) The images made with modified script but $T_{\rm sys}$ is extrapolated from 4 ATM scans. (7) The measured flux errors for the images made with original $T_{\rm sys}$ table (8) The beam size of images using original $T_{\rm sys}$ table.
Rows: a. The maximal differences for fluxes at each column.

Table 4. Flux measured for data for Band7b project (AS205A)

Dataset Label	Flux (Jy)			Meas. Err. (Jy)	Beam (")
	Tsys_orig	Tsys_extrap	Tsys_partextrap		
(1)	(2)	(3)	(4)	(5)	(6)
Band7b1	0.7699	0.759	0.7594	0.031	1.13 x 0.79
Band7b2	0.713	0.7146	0.7129	0.029	0.96 x 0.77
Band7b3*	1.001	1.005	1.003	0.043	0.68 x 0.53
Band7b4	0.747	0.7378	0.7359	0.028	1.03 x 0.78
Band7b5*	0.556	0.58	0.58	0.029	1.31 x 0.72
Band7b6	0.749	0.7534	0.7533	0.03	0.91 x 0.08
Band7b7	0.7355	0.7347	0.7348	0.029	1.01 x 0.81
Band7b8	0.768	0.737	0.7372	0.029	1.24 x 0.79
AVG.[a]	0.7471	0.7394	0.7389		
STD.[b]	0.0194	0.0143	0.0149		

Columns: (1) The label for each dataset (see Table 2). (2) Flux of the data set using the original $T_{\rm sys}$ table. (3) Flux of the data set using alternative continuous $T_{\rm sys}$ table with the linear relation fitted using all ATM scans. (4) Flux of the data set using continuous $T_{\rm sys}$ table with the linear relation fitted using part of ATM scans. (5) Measured flux errors using original $T_{\rm sys}$ table. (6) Beams of the image using original $T_{\rm sys}$ table.
Rows: a. The average value for each column. b. The standard deviation for each column.
Notes: * denotes data with unusual fluxes. Fluxes from these data are not included in the calculation of the average and standard deviation value.

Table 5. Flux measured for data for dataset Band9b (Circinus)

Dataset Label	Flux (Jy)			Meas. Err. (Jy)	Beam (")
	Tsys_orig	Tsys_extrap	Tsys_partextrap		
(1)	(2)	(3)	(4)	(5)	(6)
Band9b1	0.85	0.875	0.849	0.018	0.086 x 0.065
Band9b2	0.96	1	0.983	0.021	0.104 x 0.077
Band9b3	1.27	1.35	1.34	0.016	0.131 x 0.085
AVG.	1.027	1.075	1.057		
STD.	0.178	0.201	0.207		

Columns: (1) The label for each dataset (see Table 2). (2) Flux of the data set using the original $T_{\rm sys}$ table. (3) Flux of the data set using alternative continuous $T_{\rm sys}$ table with the linear relation fitted using all ATM scans. (4) Flux of the data set using continuous $T_{\rm sys}$ table with the linear relation fitted using part of ATM scans.
Rows: a. The average value for each column. b. The standard deviation for each column.

then the effect of bright lines in such cases will be negligible. But for very bright continuum sources such as planets, the spectrally-averaged $T_{\rm sys}$ will potentially be affected by the target brightness. However, the beam of the WVR unit on each antenna is offset from the optical axes of the receiver beams by several arc minutes (depending on the receiver band in use - see ALMA Technical handbook); this means that the WVRs are not pointing to the science target, and in general will not be affected by its strong continuum. Additionally, it has recently become apparent that the method used by ALMA to measure $T_{\rm sys}$, using off-source data along with the normalisation of the visibilities using the autocorrelation, introduces a calibration error for bright sources. The planned change is to measure Tsys on-source. Again, this should not significantly affect the continuous Tsys method, for reasons given above.

For spectral lines, an assumption is made that $T_{\rm sys}$ is mainly affected by PWV, and the correlation of Tsys with PWV uses $T_{\rm sys}$ averaged over the spectral window. This is considered reasonable for continuum and most spectral lines, but for calibration of spectral lines coincident with deep Ozone absorption (eg see Figure 12), the correlation will have a slightly different slope and intercept. In general this is considered a second-order effect; for example, a line exactly coincident with the strong O_3 peak at 428.8GHz in Fig. 12, the error in the correction of $T_{\rm sys}$ based on the PWV would be $\sim \pm$ 3%. A future improvement might be correct the data spectrally rather than using a single channel-averaged value per timestamp. However, this would make the correction table significantly larger (see Section 4.1).

An additional use of the continuous $T_{\rm sys}$ method could be to correct for the increase in $T_{\rm sys}$ due to shadowing of the antennas. On ALMA, the default is that data taken with any slight blocking of the beam from an antenna, for example by a nearby antenna or building, is flagged and removed during data reduction. In general this cannot be corrected for using the gain calibrator amplitude solution, as this is not observed at the same sky location as the target. However, if the corresponding increase in Tsys due to shadowing is measured continuously, it may be possible to calibrate out some degree of shadowing. Further investigation of this technique should be done.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we explore a new method to use continuous datastreams available from WVR monitoring to track the atmospheric opacity and hence $T_{\rm sys}$ in mm and submm data. The aim is to improve flux calibration in conditions where the sky opacity is rapidly varying, and to reduce overheads needed for frequent discrete calibration using internal loads. Here we summarize our main conclusions regarding initial tests of this method.

- There is a tight linear correlation between normalized $T_{\rm sys}$ and $T_{\rm WVR}$, with typical scatter of $\sim$ 1%. Although the exact form of the linear relation varies among different spectral windows and different data sets, we can use as few as 4 ATM scans to determine the slope and intercept of the linear relation, which suggests it is possible to significantly reduce the number of discrete $T_{\rm sys}$ measurements during observations, particualy at high frequencies. Furthermore it is not necessary to perform separate calibrations on the phase calibrator and science target, as the continuous $T_{\rm sys}$ method is able to track differences in $T_{\rm sys}$ between the two. We also try the ATM modeling and find a good agreement between the data and ATM generated correlation between $T_{\rm sys}$ and $T_{\rm WVR}$.

- We apply the continuous $T_{\rm sys}$ in calibration and find that it generally gives us more consistent fluxes for the same target. The only exception is the dataset Band9b as our new methods give higher flux uncertainties. We suspect part of reasons are due to the imperfect linear fitting of the $T_{\rm sys}$ vs $T_{\rm WVR}$ relation. Since the uv-coverages for the data sets in this project are significantly different, it is hard to confirm this scenario for this data set.

- If this method is used for sub-mm observatories such as ALMA, it can reduce the number of $T_{\rm sys}$ measurements required at high frequencies from $10 \sim 20$ down to 5 (4 $T_{\rm sys}$ measurements for the fitting and 1 bandpass $T_{\rm sys}$) or fewer. Assuming each observation block takes ~ 60 mins and each $T_{\rm sys}$ measurement takes about $30 - 40$ seconds, it has the potential to save $\sim 10\%$ of observing time for high frequency observing, which is made more valuable as the amount of time in such good conditions is limited.

ACKNOWLEDGMENTS

We thank the referee for thoughtful comments and constructive suggestions, particularly the part ATM modeling. This paper makes use of the following ALMA data:
ADS/JAO.ALMA #2015.1.00271.S
ADS/JAO.ALMA #2016.1.00744.S
ADS/JAO.ALMA #2018.1.01778.S
ADS/JAO.ALMA #E2E8.1.00003.S
ADS/JAO.ALMA #2018.1.01210.S
ADS/JAO.ALMA #2019.1.00013.S. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This research made use of Astropy,[1] a community-developed core Python package for Astronomy (Astropy Collaboration et al. 2013, 2018). HH acknowledges the support of NSERC-CREATE NTCO training program. The research of C.D.W. is supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Canada Research Chairs program.

Software: astropy (Astropy Collaboration et al. 2013, 2018), CASA (McMullin et al. 2007)

REFERENCES

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33, doi: 10.1051/0004-6361/201322068

Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123, doi: 10.3847/1538-3881/aabc4f

Bachiller, R., Carilli, C., Cox, P., et al. 2003, ALMA Science Advisory Committee (ASAC), 2003, "https://www.nrao.edu/archives/items/show/34475"

Brogan, C. 2018, Advanced Calibration Topics - I, "https://science.nrao.edu/science/meetings/2018/16th-synthesis-imaging-workshop/talks/Brogan_Adv_Cal_1.pdf"

Condon, J. J., & Ransom, S. M. 2016, Essential Radio Astronomy

Hills, R. 2004, ALMA Memo No. 495

Hills, R., Gibson, J., Richer, J., et al. 2001, ALMA memo 352

Mangum, J. 2002, ALMA Memo No. 434

—. 2017, ALMA Memo No. 602

McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell, 127

Moreno, R., & Guilloteau, S. 2002, ALMA memo 372

Pardo, J. R., Cernicharo, J., & Serabyn, E. 2001, IEEE Transactions on Antennas and Propagation, 49, 1683, doi: 10.1109/8.982447

Payne, J., Vaccari, A., Emerson, D., & Mangum, J. 2001, ALMA Construction Project Book, Chapter 3 Section 2.

Remjian, A., Biggs, A., Cortes, P. A., et al. 2019, ALMA Technical Handbook,ALMA Doc. 7.3, ver. 1.1, 2019, doi: 10.5281/zenodo.4511522

Yun, M., Bastian, T., Holdaway, M., Mangum, J., & Welch, J. 1998, ALMA Memo No. 211

APPENDIX

A. EXTRA FIGURES

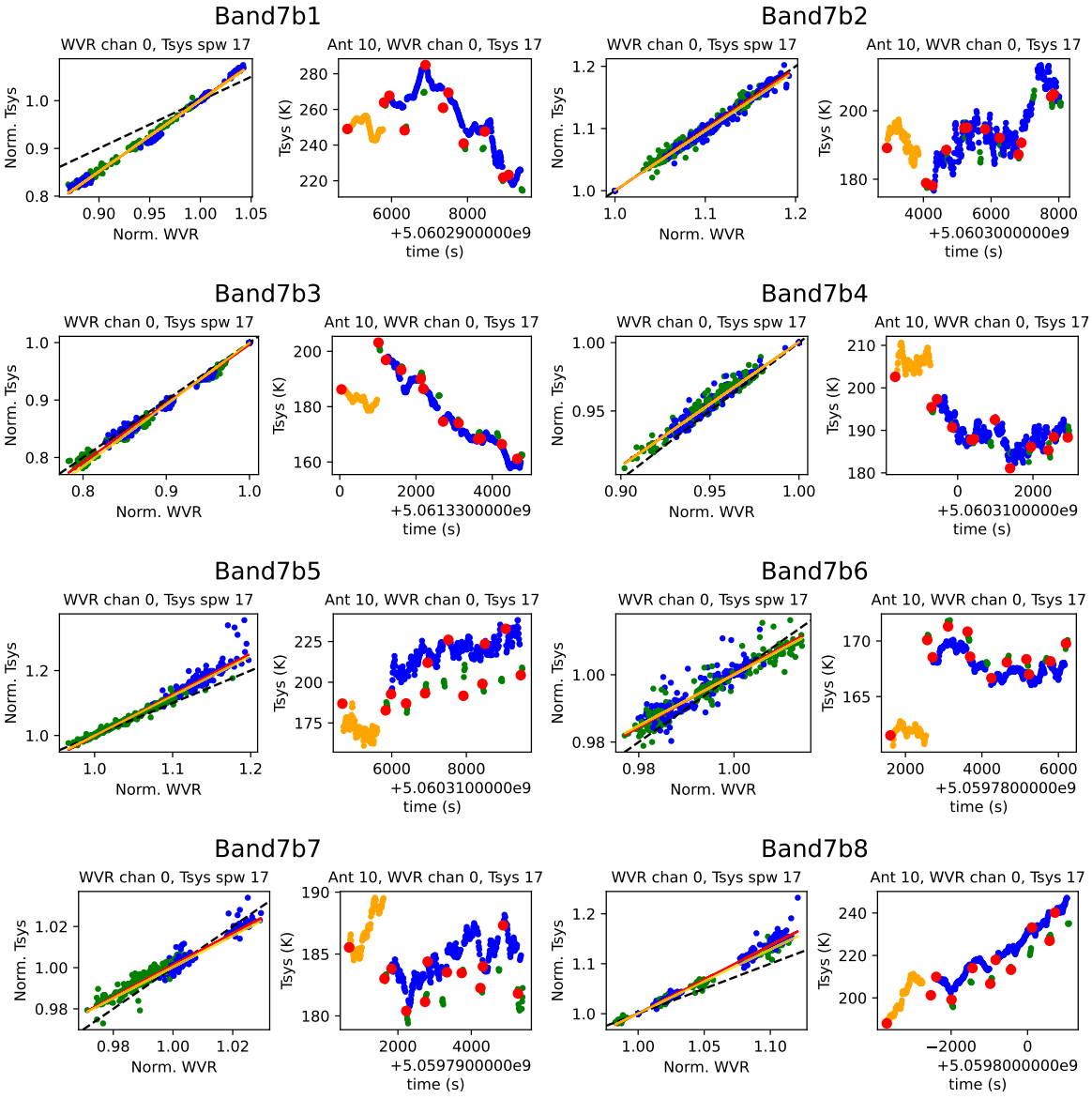

Figure 22. The summary of linear relation between $T_{\rm sys}$ and $T_{\rm WVR}$ and comparison between measured and extrapolated $T_{\rm sys}$ for measurement sets in project 2018.1.01210.S. For each measurement set, the left subplot shows the linear correlation between normalized $T_{\rm sys}$ and $T_{\rm WVR}$. The blue and green points are from science and phase scan. The dashed line shows the 1-to-1 relation. The red and golden solid line is the fitted linear relation using data from all ATM scans or just 4 ATM scans. The right plot shows the extrapolated $T_{\rm sys}$ based on the fitting relation using all ATM scans. The orange, green and blue points are extrapolated $T_{\rm sys}$ for bandpass, phase and science targets. The red points are the original measured $T_{\rm sys}$.

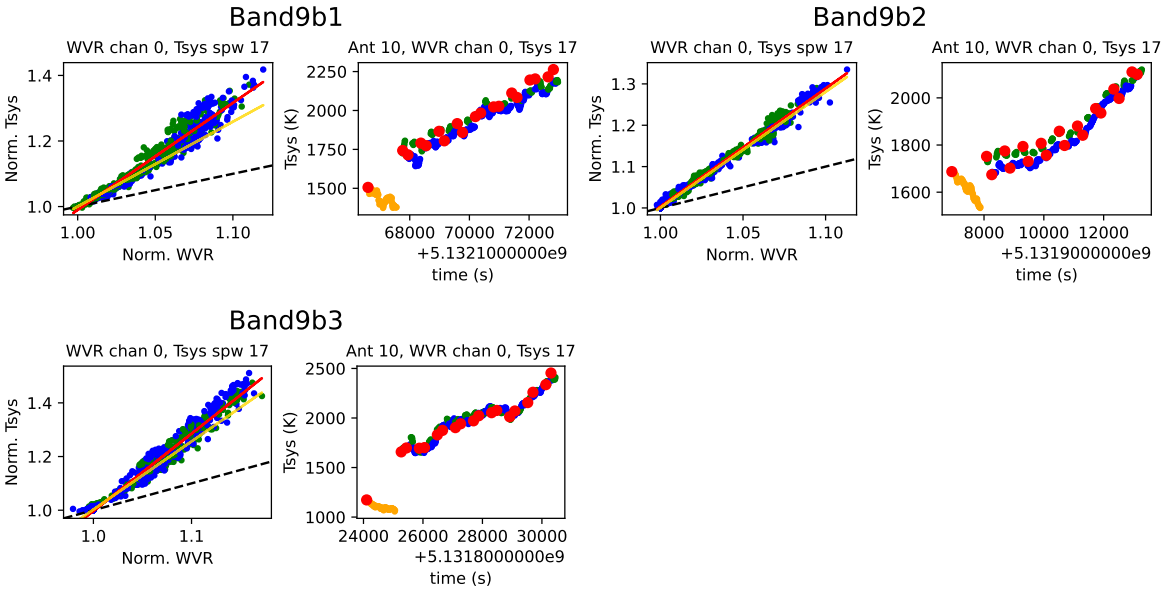

Figure 23. The summary of linear relation between $T_{\rm sys}$ and $T_{\rm WVR}$ and comparison between measured and extrapolated $T_{\rm sys}$ for measurement sets in project 2019.1.01210.S. See details in Fig. 22.

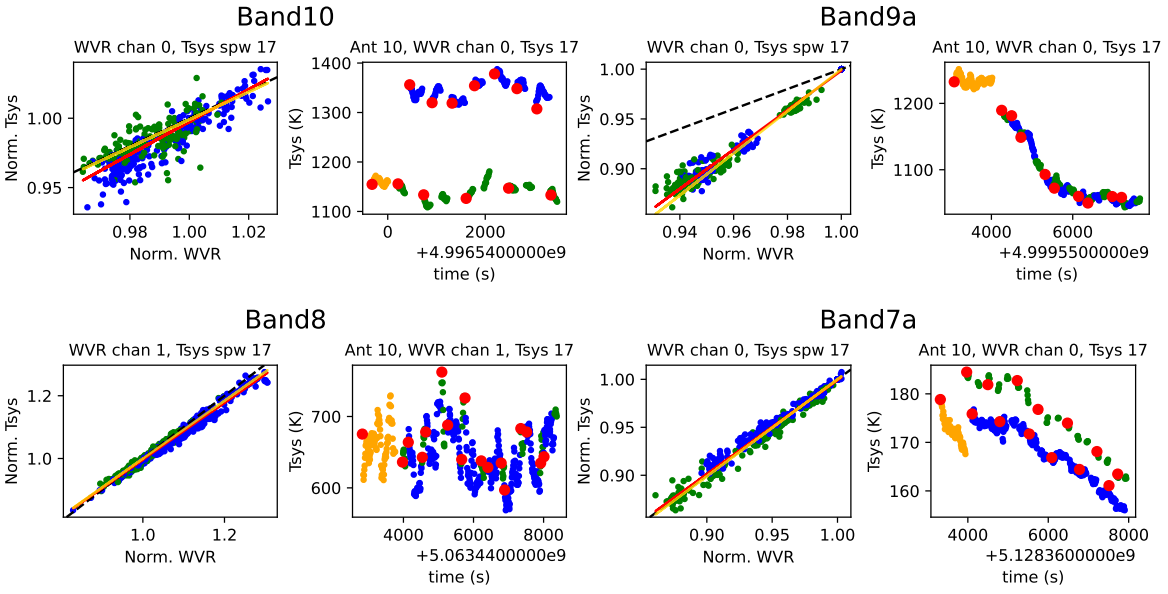

Figure 24. The summary of linear relation between $T_{\rm sys}$ and $T_{\rm WVR}$ and comparison between measured and extrapolated $T_{\rm sys}$ for the other 4 projects used in this paper with just one measurement set. See details in Fig. 22.

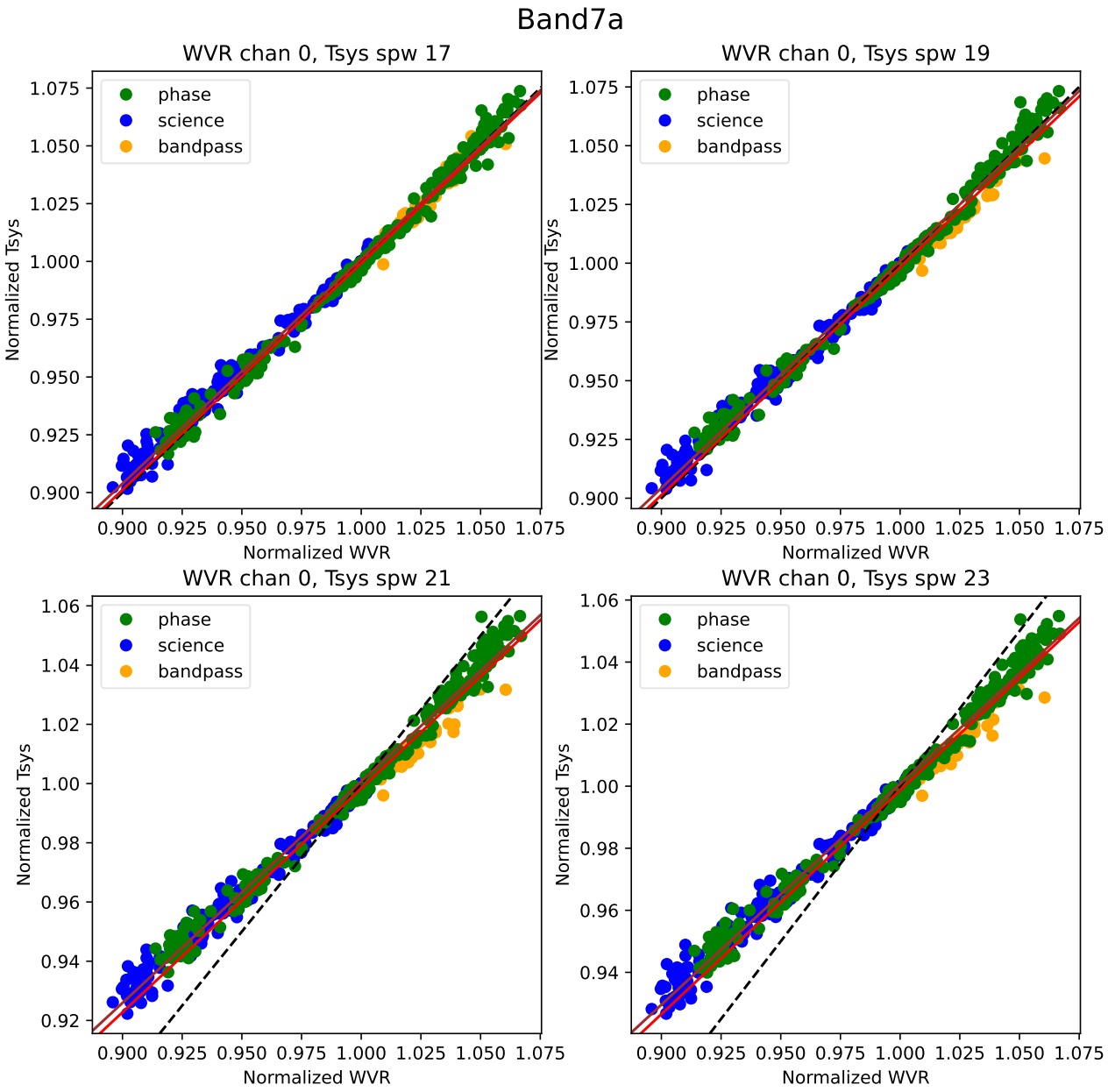

Figure 25. $T_{\rm sys}$ versus $T_{\rm WVR}$ normalized to the first science ATM scan for the dataset Band7a. The black dashed line is the 1-to-1 relation. The brown solid line is the linear fitting to the data. The red solid line is the original fitting relation to the data normalized to each type of observing target.