A refined method of the data processing for astroclimate measurements in mm-waves

Grigoriy M. Bubnov[*,1,2], Vyacheslav F. Vdovin[1,2], Peter M. Zemlyanukha[1] and Igor I. Zinchenko[1]

[1]*Institute of Applied Physics RAS, N.Novgorod, 603950, Russia*

[2]*Nizhniy Novgorod State Technical University, N.Novgorod, 603950, Russia*

*Contact: bubnov@ieee.org

***Abstract*— We are presenting an algorithm of post-processing of data obtained through the atmospheric dip method, and the properties of the filters to be used with this algorithm. The algorithm allows to make flagging of erroneous data and thus to decrease the stochastic error of the atmospheric transparency values. The algorithm includes finding and excluding the own noises of the receiver, flagging the data corresponding to broken or stratus clouds, and calculating the tau with different methods. Based on physical parameters of the atmosphere and the hardware, the algorithm helps save from 5 to 20 percent of data which otherwise are not suitable for processing.**

I. Introduction

Atmospheric propagation of terahertz waves strongly depends on the atmospheric conditions, and one of the topical issues of today is to investigate this dependence and develop techniques, instrumentation and mathematical modelling of measurement of the atmospheric transparency. Our improved post-processing method described below in this paper can be applied to data collected with the atmospheric dip method. Our method includes finding the own-noise parameters of the hardware, evaluation of the atmospheric conditions with the measurement data, and, finally, filtration of these data. One of the problems of atmospheric transparency data processing which researchers often have to deal with is gaps in data arrays which occur in partial cloudiness, and we are suggesting a way of dealing with this difficulty.

We have developed instrumentation and methods for investigations of atmospheric propagation of terahertz waves [1] and since 2012 we have gone on 7 astroclimate expeditions and explored atmospheric absorption on some 11 sites [2,3]. Our goal is to find the most appropriate place to build a radiotelescope operating in the millimetre and submillimetre wavebands. For estimating the atmospheric absorption (optical depth, tau, or Nepers) we use an assemblage which includes a radiometric system and a firmware system, the former comprising two self-contained radiometers operating in two different bands of 84-99 GHz (3 mm) and 132-148 GHz (2 mm), a rotary support, and a control. The integral absorption is determined by using the atmospheric dip method, that is, a 5-step approximation of atmospheric brightness temperature as the function of the zenith angle. The zenith angles choice is 0°, 60.48°, 76.32°, 81.36°, 84.24° and 88.56° (the latter corresponds to aiming at the horizon). Each of these angles corresponds to an integer number of steps of the stepper motor by which the mirror is driven, and to more or less equal intervals between the brightness temperatures at each of these directions. The voltage of the receiving detector (we use the negative value of this for the sake of convenience) is in linear dependence on brightness temperature; linearity holds over a wide range of brightness temperature values.

For the case of the clear sky, the experimental points can be fitted well by exponent function, and the result of approximation (the exponent index, Nepers) has a good stability over time. But for the case of partial cloudiness the fitting occurs with wide deviation against the exponent and large scattering of the points in time.

The design of the algorithm was due to the properties of our hardware and the method of the atmospheric depth measurement. The atmospheric dip method is based on measuring the difference of brightness temperature of the atmosphere's own radiation at different elevations. If the gradient of the brightness temperature is broken for some reason (cloudiness or the own noise of the receiver) this results in either fatal calculation errors (the logarithm of a negative number or division by zero) or dramatic increase in inaccuracy.

At the same time, we cannot filter or process experimental data unless such procedure is well-grounded from the physical point of view. Thus, the data processing task is formulated as follows: first, we must take away the error in the measurements

as long as it is possible, and then muffle of the remaining noises of the receiver and the atmosphere with the help of mathematical processing.

II. Algorithm

The algorithm includes several steps united by the uniform logic which is being discussed at the end of this chapter. In addition, we have annexed the Python script used at one stage of the algorithm to this paper. The primary data are output by the software controlling the physical interfaces of the MIAP-2 receivers: it gathers the measurements, drives the mirror, reads the ADC readings and records a Raw file indicating voltage of the detectors at each angle. The files are then mathematically processed.

A. Calculation of optimal integration time

Integration time of the atmospheric brightness temperature at each angle can be changed within a wide range of values each time the software controlling the physical interfaces of MIAP-2 is launched. This does not apply to the previously collected data.

As we lower the integration time for a given angle, we increase the stochastic spread of values. If we raise the integration time, the average one-cycle period will also increase, and therefore, short-time fluctuations of the atmospheric conditions will be lost. Figure 1 shows a diagram of finding the optimal integration time for a given angle.

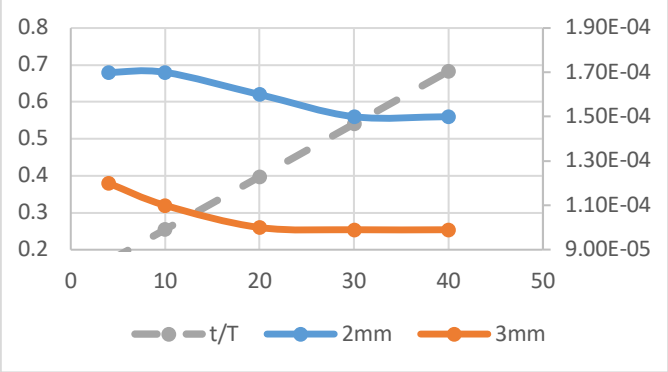

Fig. 1. A diagram showing the dependence of RMS noise of the data (2 mm and 3mm, right-hand axis) and catching of the short-time fluctuations in atmospheric conditions on the integration time at a given angle (left-hand axis).

We grant that the reference time of change of the atmospheric conditions is 7 minutes. This is a tentative parameter defined empirically based on experimental data. The grey dashed line reflects the ratio of the one cycle measurement time to the reference time of 7 minutes. Similarly to the conclusions following from the Kotelnikov theorem, we think it reasonable to grant that the optimal ratio is about 0.5, that is, the time for one-cycle measuring should be twice as little as the time of the change in the atmospheric parameters. The continuous lines show the measured sigma values (RMS) in both channels. These values obviously drop as the integration time grows. Therefore, the optimal integration time for a given angle can be defined as about 20 to 30 seconds.

B. Measuring and filtering the circuit noises

It is evident that the receivers have their own thermal noises. The correction cannot include the atmospheric parameters for the obvious reason that we cannot know them before we measure them. The filtering procedure is based on the parameters of the hardware which are known or can be measured so that we could know that we would filter out of the experimental data only the hardware-generated noises. In order to measure the hardware own noises, the quasi-optical path of the receiver is covered with the black body absorber whose temperature is stabilised with the help of a standard air conditioning device, and then the readings are recorded in the usual manner.

The resulting record of noises is then wavelet-packet decomposed into order-2 coiflets [4] with the help of the PYWT library [5]. Further, the median absolute deviation is calculated in the wavelet space, and thus the low pass filter is defined. We apply this filter to the collected observation data. An example of the filter application is shown in Figure 2.

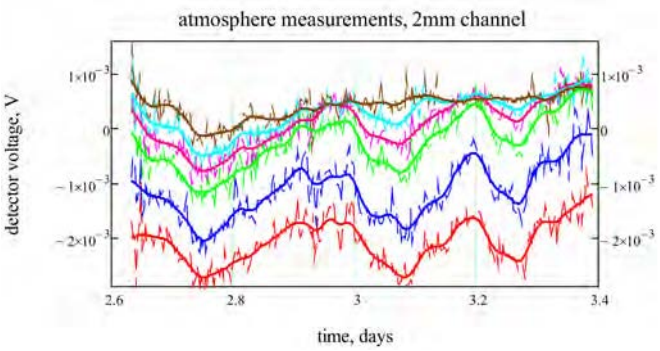

Fig. 2. Filtration of data in the 2-mm channel of the receiver. The topmost plot represents the horizon-aimed measuring, the bottom line the zenith, and the intermediate positions are between them. The thin dashed lines show the measuring, while the thick lines represent the filtered data.

The filtering parameters for the 2-mm and 3-mm channels are calculated independently. For the 3-mm channel the filtration parameters are calculated in a similar manner; we only have chosen the 2-mm channel to be presented here as a more illustrative example. The filtered data are then processed through the further stages of the algorithm.

C. Cloud correlator

When taking measurements in low partial cloudiness there is a high probability that the receiver's beam will cross a cloud and

cause the brightness temperature to rise over a relatively short time interval at the given angle. This temporally and spatially local upsurge muddles the evaluation of the optical depth and must be filtered out of the further consideration. In order to pinpoint such upsurges, we calculate the Pearson correlation for each pair of angles and then clear away everything below 0.8 from the data array.

The example in Figure 3 shows the upsurge of the brightness at the beginning of the third day (see the segment between 3.03-3.12 in Fig. 3) as the beam of the receiver was aimed at a cloud at 60.48° thus decreasing the Pearson correlation down to 0.6. In the further processing all data corresponding to the Pearson correlation below 0.8 were excluded.

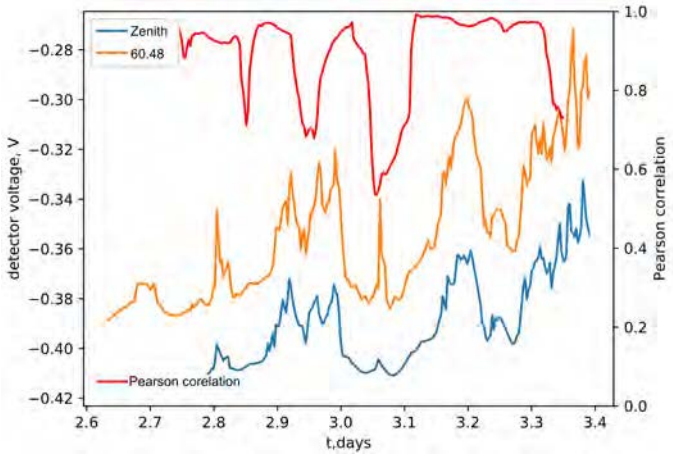

Fig. 3. An example of the Pearson correlation calculation for the zenith and 60.48° pair.

D. *Flagging of overlapping readings*

In high humidity or overcast conditions some values measured at low angles overlap the readings taken at the horizon. This happens mostly due to the dispersion of radiation in the stratus clouds; at the low angles the receiver also senses the thermal radiation of the land surface.

The algorithm tracks down converging of the receiver voltage at juxtaposed angles going over from point to point; if the convergence becomes lower than 2σ, the respective interval will be excluded from the consideration. The horizon and the zenith are reference points and never deleted.

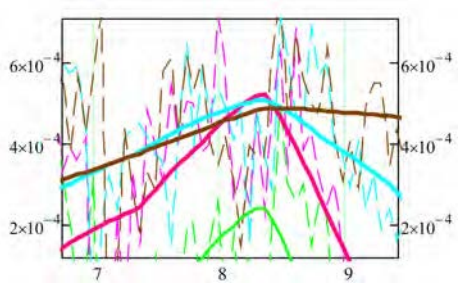

Fig. 4. An example of overlapping of voltages at angles near the horizon: voltages at 84.24° and 81.36° converge and cross the voltage taken at the horizon.

E. Calculation of tau

After the voltages of each particular angle are independently processed the data are used for calculating the optical depth (tau or Nepers). At this stage the voltages taken at different angles are fitted to Equation (1) using two methods: the least square fitting and k-Nearest Neighbor.

$$U(\theta) = U_{aver}\left(1 - e^{-\tau \cdot \sec(\theta)}\right) + U_{shift}$$

where τ is optical depth, Nep;

U_{aver} is the mean voltage corresponding to the brightness temperature of the atmosphere at the horizon;

U_{shift} – is the zero-shift voltage conditioned by the hardware properties;

U(θ) – are the measurements at particular angles.

This equation does the fitting of three parameters for each point over the time: U_{aver}, U_{shift} and τ. If for a particular moment 1, 2 or 3 angles were removed from consideration, the fitting is performed for the remaining angles. Exclusion of angles leads to the increase in the theoretical RMS of the evaluation, but not to the data loss as it previously was the case with the older algorithm.

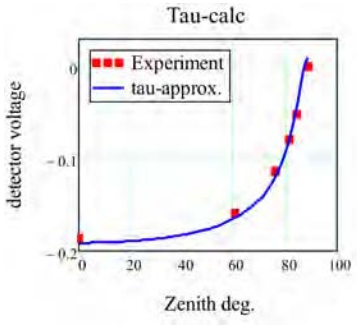

Fig. 5. Fitting of the experimental data with the exponent.

We used two methods for the calculation of tau: the least-squares fitting (LIMFIT [6], LSQ) and the k-Nearest Neighbors (SCIKIT-LEARN [7], kNN). Assuming random noise parameters and the LSQ fitting error we introduce the kNN algorithm for the tau estimation in mostly the same way as discussed in [8]. We use previous measurements and tau estimation as a training set for kNN and LSQ with the m=2 Minkowski metric. As soon as we have enough exposures to fill in neighbouring spaces we obtain RMS for tau $\sqrt{k(k-1)}$, where k is the number of neighbours. In this paper we use 300,000 modelled observations as neighbouring spaces and receive only $\sqrt{k-1}$ RMS reduction. In this case kNN is an addition to LSQ algorithm, not the separate one.

F. Algorithm as a whole

The flowchart in Figure 6 shows data flowing from one stage of the algorithm to the next one indication the data formats. The algorithm is performed in three different software environments. Thee collection of data and physical interface control are done by MIAP-2, the integration time is calculated by MathCAD, and the collected data are processed by Python.

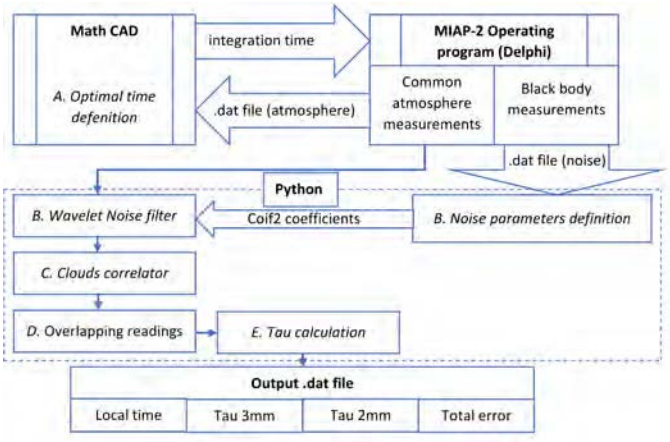

Fig. 6. The algorithm flowchart.

III. Data Processing Results

As an example, we present here data collected from the roof of one of the buildings of the IAP RAS in Nizhny Novgorod in winter time (February to March 2018). All three types of cloudiness are found in the example data: clear sky, low partial cloudiness, and sky overcast with stratum clouds. We chose the optimal integration time to be 20 seconds and recorded the own noises of the receiver applying the same integration time (Fig. 7).

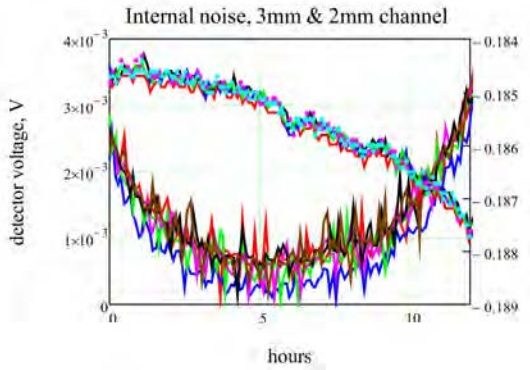

Fig. 7. Record of the hardware own noises over several hours.

The wavelets for the Coif2 filter were calculated as

2mm 0.00070858

3mm 0.00257025

The cloud correlator excluded partial cloudiness from our data arrays, for example, the one that occurred at the beginning of the third day of observations at 60.48° in the 3-mm channel. An example of the overlapping readings flagging is shown for the 2-mm channel at the end of the second day of observations (Fig. 8).

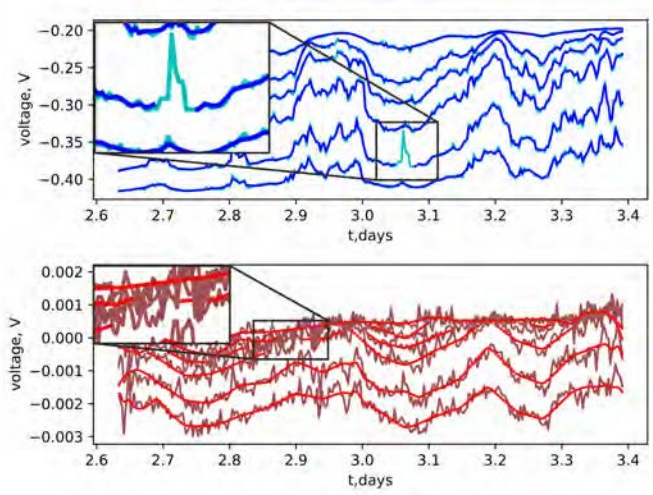

Fig. 8. Examples of algorithmic exclusion of partial cloudiness (top, 3—mm channel) and overlapping readings (bottom, 2-mm channel).

Figure 9 shows the results of the tau calculation for both channels as compared to the old method of direct calculation of an iterative equation by 5 angles. The new algorithm permitted to decrease the RMS deviations and partially to keep the data previously unsuitable for use.

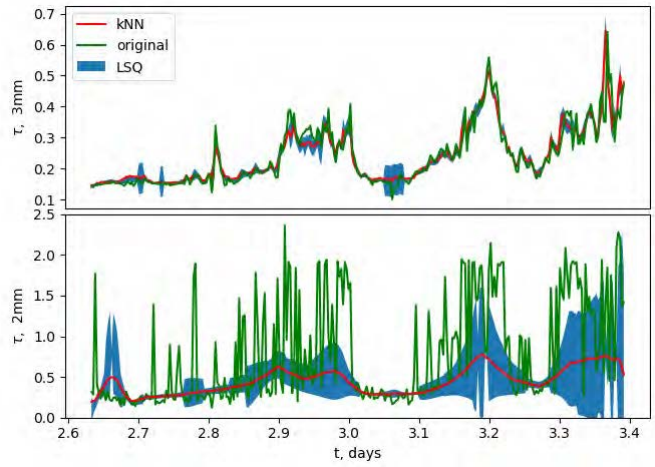

Fig. 9. The results of the algorithm as compared to the standard method.

Conclusions

In this paper we present the algorithm for collection and processing of data obtained with the atmospheric dip method. Its performance is stable and has been broken in for the

hardware we dispose of; presently, we are working on a simplified user-friendly interface. The algorithm presented here is optimal for our hardware; however, it is possible that the algorithm will require modifications to be applied to hardware operating in different wavebands.

ACKNOWLEDGMENT

This work was prepared with the financial support of the IEEE Antennas and Propagation Society Eugene F. Knott Memorial Doctoral Research Grant and by the Russian Foundation of Basic Research (RFBR) within Grant 16-32-0038716 mol_a. The authors would like to thank Dmitri Sofronov for the help in editorship.

REFERENCES

[1] V. I. Nosov et al., "A dual-wave atmosphere transparency radiometer of the millimeter wave range", Electronics and Radio Engineering Instruments and Experimental Techniques, Volume 59, Issue 3, pp 374-380, May 2016.

[2] Grigoriy M. Bubnov et al., "Searching for New Sites for THz Observations in Eurasia", IEEE Tr. on TST, Volume: 5, Issue: 1, pp.64-73, 2015.

[3] G. M. Bubnov et al., "The results of the astroclimate investigations in short-millimeter wave on the Suffa plateau", Radiophysics and Quantum Electronics, Vol 59, Issue 8-9, pp.852-861, 2016.

[4] Wavelet browser by PYWAVELETS [Online]. Available: http://wavelets.pybytes.com/wavelet/coif2/

[5] Lee G, Wasilewski F, Gommers R, Wohlfahrt K, O'Leary A, Nahrstaedt H, and Contributors, "PyWavelets - Wavelet Transforms in Python", 2006, [Online], https://github.com/PyWavelets/pywt .

[6] Newville, M., Stensitzki, T., Allen, D. B., & Ingargiola, A. "LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python" 2014, September 21, Zenodo. http://doi.org/10.5281/zenodo.11813

[7] Pedregosa et al., "Scikit-learn: Machine Learning in Python", JMLR 12, pp. 2825-2830, 2011.

[8] Zemlyanukha, P.M., Zinchenko, I.I., Salii, S.V. et al. "The Spatial–Kinematic Structure of the Region of Massive Star Formation S255N on Various Scales" Astron. Rep. (2018) 62: 326. DOI: S1063772918050074.