A Statistical Learning Approach to Chang’E Microwave Radiometer Data Calibration

K.T. Tsang, S. Fu, J. Gu, and M. Zhou
Statistics Program, Division of Science & Technology
United International College, BNU-HKBU
Zhuhai, China

K.L. Chan, and Y.C. Zheng
Center for Space Science Research
Hong Kong University of Science and Technology
Hong Kong, China

Abstract—We have proposed a statistical learning model which accounts for the effect due to the Chinese Chang’E (CE) cold space calibration antenna may have been contaminated by thermal microwave radiation from the lunar surface because its field of view overlapped with the lunar surface instead of aiming just at the cold space. This effect can corrupt the simple two-point calibration procedure and lead to the discrepancy between the available microwave radiometer Level 2C data from CE1 & 2. An algorithm to implement the statistical learning model is developed. Recalibration of the Level 2C radiometer data from localized locations (around the Apollo 17 lunar mission landing site on the lunar surface) by regression analysis based on our statistical learning calibration model successfully removes the discrepancy between the two sets of microwave brightness temperature data from CE1 and 2. Residual analysis on the combined dataset shows that the residuals are normally distributed with a mean value equals to zero. This sophisticated statistical learning algorithm developed here may be considered as a recalibration effort to remove the systematic inconsistency between the two microwave brightness temperature datasets. The combined and enlarged CE radiometer dataset may potentially increase our understanding of the physical reality hidden behind.

Keywords—Statistical; Learning; modelling; microwave; brightness temperature; remote sensing; calibration

I. INTRODUCTION

Thermal environment on the lunar surface is an important consideration that affects the planning of human or robotic activities for future lunar exploration. Historically, microwave remote sensing, which is based on the use of radiometers to measure the thermal microwave radiation emitted from the target, has been the preferred method to determine lunar surface temperature due to its low power requirement.

Recently, the multi-channel (at 3.0, 7.8, 19.35 and 37GHz) microwave radiometer (MRM) technology was employed by the Chinese Chang’E lunar orbiters [1,2] to measure lunar thermal microwave emissions continuously in space and time. In two separate missions, Chang’E-1 (CE1) and Chang’E-2 (CE2), launched on Oct 24, 2007 and Oct 1, 2010 respectively, acquired large amount of raw data, which were calibrated [3,4] and subsequently distributed to the scientific community as the Level 2C MRM data in PDS format (http://moon.bao.ac.cn/). The radiometer measurements were converted to brightness temperatures (TB), with a two-point calibration procedure, which is the most common practice in microwave remote sensing [5].

The multi-channel CE-MRM is consisted of four observation parabolic antenna dishes pointing to the lunar surface, four calibration parabolic antenna dishes pointing to the cold space, one integrated unit for all microwave receivers, sets of waveguides and high frequency cables linking the antenna and microwave receiver, and a united data management unit. The observation antenna dishes and the calibration antenna dishes are mounted on two different sides of the CE orbiter body which are adjacent and perpendicular to each other. The actual mounting configuration of the two sets of the MRM antenna dishes on the CE orbiter body is shown in Figure 1, with the observation antenna dishes pointing to the lunar surface and the calibration antenna dishes aligning along the flight path and facing forward.

While the CE calibration antenna is supposed to point to the cold space and measuring data to be used as the cold reference point in the two-point calibration method [3,4] it actually received some thermal microwave radiation from the lunar surface due to the finite size of its field of view (FOV). Thus the “cold” reference point is in reality not exactly cold and this effect affects the quality of the calibration. In fact, the contamination of thermal radiation from the Earth to the cold calibration antenna is well known to the Earth remote-sensing community [5].

In this work, motivated by the difference between the available microwave radiometer level 2C data from CE1 and CE2, we have re-examined the calibration procedure of the CE MRM data. To correct for the inadvertently picked up thermal contamination from the lunar surface, an improved calibration model is proposed to rectify the situation. Recalibration of the level 2C data by statistical regression analysis based on our calibration model and algorithm successfully removed the discrepancy in a concrete example, involving data obtained around the Apollo 17 lunar mission landing site on the lunar surface.

In the following, a detail description of the original two-point calibration method [3,4] is reviewed in Section II. The new calibration model based on the statistical learning approach we proposed is discussed in detail in Section III. An algorithm to implement the statistical learning approach is given in Section IV. The results of this algorithm applied to a
specific group of data measured around the Apollo 17 lunar mission landing site (20.19°N 30.77°E) with roughly a total of 1100 records of data spanning the diurnal cycle is also shown in Section IV. Finally, the conclusion and implication of our results is discussed in Section V.

II. TWO-POINT CALIBRATION MODEL FOR CE MRM

The CE1 MRM was calibrated onboard periodically, every 11.6s to assure its reliability and accuracy, using a two-point calibration method [3]. A calibrating heat source of known temperature onboard provided the high temperature reference point, and the cold/deep space cosmic microwave background served as the low temperature reference point. Judging from the quality of TB maps [6-8] produced from the CE1 data and the lunar information retrieved, it is tempting to conclude that the instrument was functioning normally during most of its life-span in orbit and the subsequent calibration procedure was successful as expected. As the CE2 MRM is an exact copy of that of CE1, the CE2 MRM data were calibrated the same way.

The calibration procedure carried out by Wang et al. [3] consisted of a linear part and a nonlinear correction. For simplicity, we adopt the linear calibration equation for CE1 MRM data [3]:

\[ T_{\text{data}} - T_c = \frac{(V_{\text{OUT}} - V_c)}{(V_{\text{HI}} - V_c)} (T_{\text{HI}} - T_c). \] (1)

In Eq. (1), \( V_c \) and \( V_{\text{HI}} \) are the output voltages (the reference points in the two-point calibration method) associated with the cold calibration antenna and the heat source, respectively, with \( T_c \) and \( T_{\text{HI}} \) the corresponding brightness temperatures. Using Eq. (1), the output voltage measured by the observation antenna, \( V_{\text{OUT}} \), is converted to a lunar surface brightness temperature, \( T_{\text{data}} \), available to lunar scientists as the Level 2C MRM data in PDS format.

If the calibration antenna is indeed pointed to the cold space, then \( T_c = 2.7K \) and \( T_{\text{data}} \) is the true lunar surface brightness temperature, \( T_b \). Since the calibration antenna is contaminated by thermal radiation from the lunar surface, \( T_{\text{data}} \) obtained from Eq. (1) is different from the actual surface brightness temperature, \( T_b \). It is easy to show that \( T_b \) is related to \( T_{\text{data}} \) by a linear relationship such that when \( T_c = 2.7K \) we should recover \( T_b = T_{\text{data}} \) exactly:

\[ T_b = T_{\text{data}} + (T_c - 2.7) (T_{\text{HI}} - T_{\text{data}})/(T_{\text{HI}} - 2.7). \] (2)

From the previous discussion, we have obviously \( T_c > 2.7 \) due to the (previously ignored) contribution to \( T_c \) from the swath of lunar surface area inside the FOV of the calibration antenna. This can be expressed as \( T_c - 2.7 = T_{\text{edge}} \) where \( T_{\text{edge}} \) is a measure of the extra microwave energy received by the calibration antenna from the small swath of lunar surface within its FOV. In general, it is difficult to account for \( T_{\text{edge}} \) since it is the sum of all radiation contributed from the entire swath of lunar surface within the FOV of the calibration antenna and so depends on the spatial distribution of \( T_b \).

Ignoring the contribution of \( T_{\text{edge}} \) in previous calibration efforts [3,4] led to difference between CE1 and CE2 MRM data that has long puzzled many lunar researchers. An example of this discrepancy between the available Level 2C CE-MRM data is shown in Figure 2, where we have plotted all the Level 2C TB data around the landing site of the Apollo 17 lunar mission (20.19°N 30.77°E) from the CE1 (blue for 37GHz, red for 19.35GHz) and CE2 (green for 37GHz, magenta for 19.35GHz) MRMs versus the local time variable known as the “hour angle”, which equals \(-\pi/2\) at dawn and \(3\pi/2\) immediately before dawn. These are roughly a total of 1100 records of data from that area in a diurnal cycle within a square of one degree in latitude and longitude. We note that in general CE2 data distributed more uniformly over the local time compared with the CE2 data. While the 19.35GHz TB data from CE1 & CE2 pretty much agree with each other, the 37GHz TB data show significant divergence especially after sunset.

III. STATISTICAL LEARNING MODEL FOR CALIBRATION

In this work, we propose a simple model to account for the effect of \( T_{\text{edge}} \) by assuming \( T_{\text{edge}} = \alpha T_{\text{data}} \) so that Eq. (2) can be rewritten as

\[ T_b = T_{\text{data}} + \alpha T_{\text{data}} (T_{\text{HI}} - T_{\text{data}})/(T_{\text{HI}} - 2.7). \] (3)

Assuming the brightness temperature of the heat source is known and \( T_{\text{HI}} = 300K \) [4] then \( \alpha \), which depends on the microwave frequency among other factors such as the orbital characteristics, can be regarded as a parameter to be determined by statistical learning method using the two set of Level 2C data from CE1 and CE2 combined to give the same \( T_b \) for each frequency channel. Implicit in this approach, we assume the difference in the Level 2C data from CE1 & CE2 we observe in Figure 2 is due to the fact that \( \alpha \) has not been properly accounted for in the original calibration procedure of Wang et al. [3] (where it was assumed to be zero). With \( \alpha \) correctly accounted for in Eq. (3) both set of data from CE1 & CE2 should converge to the same diurnal behavior in a statistical sense.

Physically embedded in the parameter \( \alpha \) in Eq. (3) are all the thermal radiation contributions from the entire swath of lunar surface within the FOV of the cold calibration antenna. Even if the spatial variation of TB is known \( \alpha \) would be difficult to calculate from first principle since it depends on physical parameters such as emissivity and reflectance, as well as other geological and topographical features of the lunar surface. Hence with just a single set of MRM data alone, either CE1 or CE2, it would be very difficult to determine the parameter \( \alpha \). However, with both set of MRM data from CE1 and CE2, we can exploit them as training data to determine \( \alpha \) by statistical learning techniques. In other words, let the data tell us what the best value of \( \alpha \) should be so that TB from CE1 and CE2 will regress to the same function of the local time.

In a previous work [9], initial attempt to apply the statistical learning approach to re-calibrate the Level 2C MRM data from CE1 and CE2 has been reported with significant improvement in the consistency between the CE1 and CE2 measurements. The major result of [9] is summarized in Figure 3, where the difference, \( T_b (\text{CE1}) - T_b (\text{CE2}) \), for the 37GHz frequency channel before (orange, original level 2C data) and after (black,
AN IMPROVED LEARNING ALGORITHM AND RESULTS

In this work, a more sophisticated statistical learning algorithm is developed to remove the limitation discussed above so that all data from the two datasets are fully utilized. To demonstrate this new algorithm, in this work we apply it to the 37GHz channel data selected in Figure 2, from the landing site of the Apollo 17.

The algorithm of the learning process is to minimize an objective function obtained by the following steps:

1. Select $a_1$ and $a_2$ from the interval [0,1] to transform the Level 2C 37GHz data from CE1 and CE2 according to Eq. (3) to two new sets $T_{37}(a_1)$ and $T_{37}(a_2)$;

2. Fit $T_{37}(a_2)$ by a polynomial in the local time variable, for example, with our data used below an 8th degree polynomial roughly gives a good fit; denote the estimate of the standard deviation of the error of the fit as $\delta_2$;

3. Using the fit obtained in Step 2 to calculate the root mean square error between $T_{37}(a_1)$ and the polynomial fit; denote it as $\delta_1$;

4. Sum of $\delta_1$ and $\delta_2$ is the target function.

The reason why $T_{37}(a_2)$ instead of $T_{37}(a_1)$ is fitted in the algorithm is due to the fact that the CE2 data have more uniform coverage in time so that the fitted polynomial will be a more accurate model of the diurnal variation. The target function so obtained depends on $a_1$ and $a_2$. Searching in the $(a_1, a_2)$ space to look for a minimum in the unit square [0,1]×[0,1] will give us the optimum values for $a_1$ and $a_2$.

Following the algorithm above using the 37GHz channel data selected in Figure 2, an intermediate step in the process of searching for the optimum values for $a_1$ and $a_2$ such that TB37 from CE1 and CE2 will regress to the same function of the local time is shown in Figure 4, where $a_1=0.05$ and $a_2=0.24$ is assumed. The CE2 TB37 (green) data after the transformation and its 6th degree polynomial fit (red), together with the CE1 TB37 (blue) data after the transformation are shown as function of the local time variable hour angle. The total error in the residual (sum of $\delta_1$ and $\delta_2$) is roughly 1.9K. This choice of $a_1$ and $a_2$ is not optimized yet, but it is much better than Fig. 2 where $a_1=0.0$ $a_2=0.0$, and is very close to the optimal value. Note that this choice of $a_1$ and $a_2$ pushes the CE1 TB37 data close to those from CE2 compared with the situation shown in Fig. 2.

Result from standard optimization routine gives a local minimum right next to $a_1=0.05$ and $a_2=0.24$. Residual analysis with respect to the combined set of optimized TB37 data from CE1 & CE2 based on the 6th degree polynomial fit are shown in Fig. 5 and Fig. 6. In Fig. 5, the residuals of the combined and optimized TB37 dataset from CE1 & CE2 measured around the landing site of the Apollo 17 lunar mission are plotted as function of the hour angle. In Fig. 6, the qq-plot (with respect to a standard normal distribution) of the residuals of the combined and optimized TB37 dataset from CE1 & CE2 measured around the landing site of the Apollo 17 lunar mission is shown. The vertical scale is that of the residuals of TB37 in K. This qq-plot confirms that the residuals of TB37 are normally distributed with mean value of 0.0 and a standard deviation of 1.8K, justifying the assumption that the combined TB37 dataset regressed to a 6th degree polynomial.

What we described above constitutes the basic learning algorithm and the regression analysis on the combined and optimized TB37 dataset. More sophisticated variations of the algorithm will likely improve the result.

DISCUSSIONS

In this work, a statistical learning model [9] proposed earlier, which accounts for the effect due to the Chang‘E cold space calibration antenna may have been contaminated by thermal microwave radiation from the lunar surface, is implemented with an improved algorithm that makes full utilization of all available CE1 and CE2 MRM data. As an example, this algorithm is applied to a specific group of data around the Apollo 17 lunar mission landing site (20.19°N, 30.77°E) with roughly a total of 1100 records of data spanning the diurnal cycle. Two parameters, $a_1$ and $a_2$, which characterize the degree of thermal contamination of the cold space calibration antennas on CE1 and CE2, are needed in the recalculation of the TB37 MRM data with our statistical learning model. Our learning algorithm successfully locates the optimal values of $a_1$ and $a_2$ that remove the inconsistency between the CE1 and CE2 TB37 MRM data down to the intrinsic statistical noise level of the MRM data. Standard optimization routine gives a local minimum of the discrepancy between the two sets of data near $a_1=0.05$ and $a_2=0.24$ in this example, which implies the degree of thermal contamination of CE1 data is less severe compared with that of CE2. This agrees well with our expectation, since CE2 had a lower orbit and so naturally its cold space calibration antenna was affected more by the thermal contamination from the lunar surface.

The example studied here demonstrates the feasibility of a statistical learning approach to the Chang‘E MRM data calibration problem using a simple model with only two parameters to be determined by the learning process. We are planning to expand the recalibration process to other frequency channels and locations on the lunar surface to study the frequency dependence and global variability of $a_1$ and $a_2$. It is conceivable that they may depend on frequency, local terrain and other physical conditions.
More elaborated model with more adjustable parameters may be needed in order to increase our understanding of the physical reality hidden behind the combined CE MRM dataset.

FIGURES

Figure:1

Figure:2

Figure:3

Figure:4

Figure:5

Figure:6
FIGURE CAPTIONS

Figure 1. The actual location of the observation and calibration antenna dishes on the CE orbiter body. The calibration antenna dishes are mounted on an X plane perpendicular to the flight path and facing forward. Observation antenna dishes are mounted on a Z plane pointing to the lunar surface.

Figure 2. Level 2C TB data from the CE1 (blue for 37GHz, red for 19.35GHz) and CE2 (green for 37GHz, magenta for 19.35GHz) MRMs versus the local time variable known as the “hour angle”, which equals -π/2 at dawn and 3π/2 immediately before dawn. A total of 540 records of data within a square of one degree in latitude and longitude around the landing site of the Apollo 17 lunar mission (20.19°N 30.77°E) in a diurnal cycle are shown. While the 19.35GHz TB data from CE1 & CE2 pretty much agree with each other, the 37GHz TB data from CE1 & CE2 show significant divergence especially after sunset.

Figure 3. The difference, TB (CE1) - TB (CE2), for the 37GHz frequency channel before (orange, original level 2C data) and after (black, actual TB) the recalibration for each narrow latitude band with a width of 2°, is shown as function of latitude. Before the recalibration, the original level 2C data from CE1 is on the average about 15K higher than those from CE2. After the recalibration, the mean value of TB (CE1) - TB (CE2) is approximately zero.

Figure 4. An intermediate step in the process of searching for the optimum values for α1 and α2 such that TB37 from CE1 and CE2 will regress to the same function of the local time is shown in this plot, where α1=0.05 α2=0.24 is assumed. Here the CE2 TB37 (green) data after the transformation and its 6th degree polynomial fit (red), together with the CE1 TB37 (blue) data after the transformation are shown as function of the hour angle. The total error in the residual is about 1.9K. This choice of α1 and α2 is not optimized yet, but it is very close to the optimal value and much better than the situation shown in Fig. 2 where α1=0.0 and α2=0.0.

Figure 5. The residuals of the combined and optimized TB37 dataset from CE1 & CE2 measured around the landing site of the Apollo 17 lunar mission are plotted as function of the hour angle.

Figure 6. The q-q plot (with respect to a standard normal distribution) of the residuals of the combined and optimized TB37 dataset from CE1 & CE2 measured around the landing site of the Apollo 17 lunar mission is shown. The vertical scale is that of the residuals of TB37 in K. This q-q plot shows that the residuals of TB37 is normally distributed with a mean about 0 and standard deviation about 2K.

ACKNOWLEDGMENT

This work was supported by internal research grant R201313 from United International College, Beijing Normal University-Hong Kong Baptist University, Zhuhai, China, and by the NSFC program (40904051), CAS program (XDA04071900), Hong Kong Scholar Program. The CE data was supported by NAOC.

REFERENCES