RFI Mitigation for the Parkes Galactic All-Sky Survey (GASS)

Peter Kalberla

Argelander-Institut für Astronomie, Auf dem Hügel 71, D-53121 Bonn, Germany

E-mail: `pkalberla@astro.uni-bonn.de`

The GASS is a survey of Galactic atomic hydrogen (HI) emission in the southern sky observed with the Parkes 64-m Radio Telescope[*]. With a sensitivity of 60 mK for a channel width of 1 km/s the GASS is the most sensitive and most accurate survey of the Galactic HI emission in the southern sky. We discuss RFI mitigation strategies that have been applied during the data reduction. Most of the RFI could be cleaned by using prior information on the HI distribution as well as statistical methods based on median filtering. Narrow line RFI spikes have been flagged during the first steps of the data processing. Most of these lines were found to be constant over long periods of time, such data were replaced by interpolating profiles from the Leiden/Argentine/Bonn (LAB) survey [5]. Remaining RFI was searched for at any position by a statistical comparison of all observations within a distance of $0.^{\circ}1$. The median and mean of the line emission was calculated. In cases of significant deviations between both it was checked in addition whether the associated rms fluctuations exceeded the typical scatter by a factor of 3. Such data were replaced by the median, which is found to be least biased by RFI and other artifacts. The median estimator was found to be equivalent to the mean, which was obtained after rejecting outliers.

---

[*]The Parkes Radio Telescope is part of the Australia Telescope which is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO

1. Survey description

The Paries Galactic All-Sky Survey (GASS) is a survey of Galactic atomic hydrogen (H I) emission in the southern sky for declinations $\delta < 1°$ observed with the Parkes 64-m Radio Telescope. The first data release was published by McClure-Griffiths [5]. The GASS is the most sensitive, highest angular resolution large scale survey of Galactic H I emission ever made in the southern sky. The intrinsic angular resolution of the data is 14′.4 (FWHM). The velocity resolution is 1.0 km s^{-1} and the useful bandpass covers a velocity range $|v_{lsr}| < 468$ km s^{-1}.

2. Data reduction

Observations and initial steps in the data reduction are described by McClure-Griffiths [5]. For the second data release ([4]) the stray radiation from the antenna diagram was calculated and subtracted after an appropriate calibration of the observations. This correction may affect profile wings, it is therefore mandatory to apply the correction for instrumental baselines after the stray radiation was eliminated. LAB data have been used for an initial estimate of the emission-free part of the profile. The instrumental baseline was fitted iteratively by a 9th order polynomial. Channels flagged previously as suspicious for RFI were disregarded in this step.

After the initial steps in the data reduction all currently known systematical influences of the instrumentation should have been taken into account. The data base has to be self-consistent. Three main criteria for consistency are available: 1) the derived H I distribution needs to match previous observations (e.g. the LAB data), 2) except for absorption caused by strong continuum sources no negative signals should exist, and 3) data observed at different seasons should not deviate from each other. Remaining inconsistencies have in the first instance to be addressed to RFI. In the following we describe our search for such problems and corrections that have been applied to the data base.

3. RFI detection

3.1 Flags set by *Livedata*

Easily detectable strong narrow-line interference was flagged during the basic data reduction with *Livedata* [5]. This kind of RFI occurs typically at fixed topocentric frequencies. As the telescope scans across the sky the RFI moves in successive channels in the LSRK frame. Most of these features extend over large angular distances and affect fields that have been observed independently at different seasons (Fig. 1). This behavior implies that the RFI remained approximately constant for long periods in time, affecting most of the data.

This kind of strong RFI has been eliminated by replacing the flagged signal by interpolated data from the LAB survey. Alternatively we have tried to interpolate the observed GASS data by using the nearest ten good data points. Using LAB data resulted in all cases in much cleaner maps. About 0.1% of all data have been flagged by *Livedata* and were replaced by interpolated data in this first step. In addition, all profiles with more then 30 flagged channels were completely discarded. Figure 1 shows an example of this kind of extended RFI (left) in comparison with the cleaned image (right). Most of the RFI is outside the band, causing therefore a negative signal.
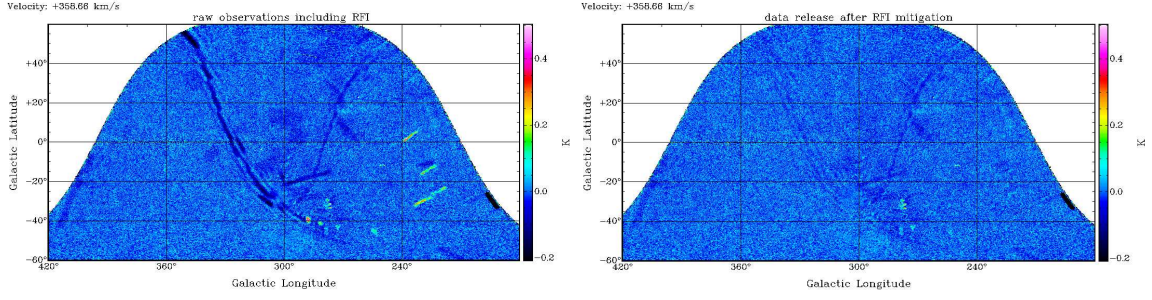
Figure 1: Typical example for extended interference patterns that were flagged by *Livedata* in the first stage of the data reduction. Left: as observed, right: after RFI removal using LAB data. The intensity scale is linear for $-0.2 < T < 0.5$ K. Some weak structures remain but are mostly at $|T| < 40$ mK

3.2 Median filter RFI excision

Even after removal of the channels flagged in the initial data reduction, numerous instances of RFI remain in the data. Some appear as point sources, but the most obvious show "footprints" of the 13 beam system that move on the sky in scanning direction (Fig. 2). As before, the motion in successive velocity channels is caused by changing LSR projections of the fixed topocentric RFI frequencies.

To search for RFI we used for each observed position a statistical comparison of all data obtained within an angle $\varepsilon = 6'$. This angle allows a comparison of typically 40 individual spectra, a number that is sufficiently robust for a statistical analysis without degrading the spatial resolution too much. Using such a sample we calculate the mean for each channel, the corresponding standard deviation, and the median, disregarding data flagged either by *Livedata* or our own post-processing. The average rms noise level $T_{RMS} = \sigma(T)$ and its one sigma deviation $\sigma(T_{RMS})$ from the average was determined for the low level emission part of the profile ($T_B < 0.5$K).

We have two tests for RFI. In the first, channels with rms deviations $\Delta T_{RMS} > 3\sigma(T_{RMS})$ were considered as highly suspicious for RFI contamination and investigated in more detail. We use as a second indication of RFI circumstances where the mean and median differ from each other by more than one standard deviation.

To eliminate outliers for the channels that were selected in both of the two tests, we determined if individually observed brightness temperatures deviate from the median by more than two times the standard deviation. In the case of a purely random distribution this would exclude 5% of the good data, and we consider this an acceptable cost to eliminate true RFI. We flag outliers and repeat, calculating new mean and median estimators and once more excluding any outliers.

Our two sigma limit may be compared with the more rigorous method for the elimination of suspect data proposed by [6], which is based on probability theory. Accordingly a rejection of a data point that deviates from the mean at a two sigma level is justified if there are 13 independent observed data values. For $\varepsilon = 6'$ we have on average 40 individual profiles and a two sigma limit is appropriate if about 10% of the data are suspect. This condition is valid for many cases but we find also situations where 20% of the measurements are outliers. In this case a cutoff at 1.6 $\sigma(T_{RMS})$ would be more appropriate according to Peirce [6]. We deviate from Peirce's criterion in two ways; to simplify the calculations we use a fixed 2 $\sigma(T_{RMS})$ cutoff and we define outliers by their

deviations from the median and not, as originally proposed by Peirce, from the mean. The latter is essential since outliers caused by RFI can be found at ten to hundred times the standard deviation, a circumstance that was not considered by Peirce; in such cases the mean is usually seriously affected.

To replace values outside the two sigma range, the best available estimate is the median value scaled by the proper beam weighting function, for details we refer to the extended discussion by [1]. We replace the previously selected data points by the weighted median and set a flag to distinguish medians from observed values. A fraction $\sim 5 \cdot 10^{-4}$ of the data is affected by RFI and replaced by the weighted median.

One of the criteria that we used to detect RFI was the fact that mean and median should not differ by more than one standard deviation. We applied this criterion also to test whether the median estimator is consistent with the mean of all data after exclusion of the outliers. Both methods were found to produce nearly identical results. An example of RFI in scanning direction and the result after application of our median filer is given in Fig. 2.
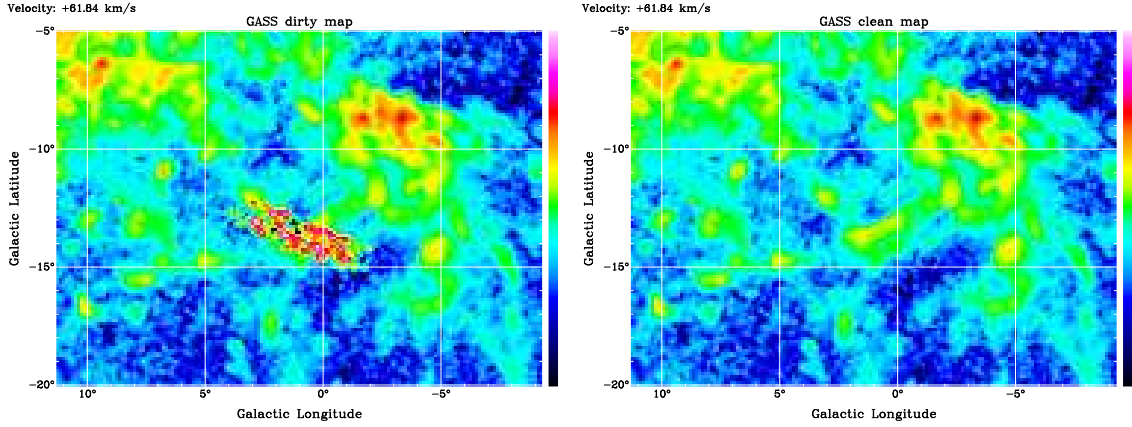
Figure 2: HI emission at $v_{lsr} = 61.8$ km s^{-1}. Data heavily affected by RFI in scanning direction of the telescope (left) are compared with the clean map (right). The intensity scale for the range $-.04 < T < 20$ K is logarithmic, emphasizing low level emission. Yellow colors correspond to a brightness temperature of ~ 1 K.

Figure 3 shows spectral details for one position in this field. Displayed are the mean (red), the rms scatter (green), and the median (blue) derived for all profiles that are closer than $\varepsilon < 6'$ to the position $l = 0°$, $b = -13°.5$. Data in the left plot suffer from strong interference at $v_{lsr} \sim 157$ km s^{-1}, $55 < v_{lsr} < 58$ km s^{-1}, and at $v_{lsr} > 400$ km s^{-1}. Only a part of the data was flagged by *Livedata* and replaced by values from the LAB. Despite all defects the median (blue) is essentially unaffected by RFI. The data after RFI elimination are shown in the plot at the right side. Profiles for mean and median are within the noise identical, the rms is close to the expected value of 0.4 K and shows no strong enhancements. We find some fluctuations close to the line emission at $v_{lsr} \sim 0$ km s^{-1}. This may have been caused by genuine fluctuations in the H I line emission. To avoid a possible degradation of such fluctuations, no automatic RFI cleaning was applied for $T_B > 0.5$K. These two plots demonstrate how greatly the rms is affected by RFI, and how deviations between median and mean are obvious. Our RFI filter is triggered by such deviations.
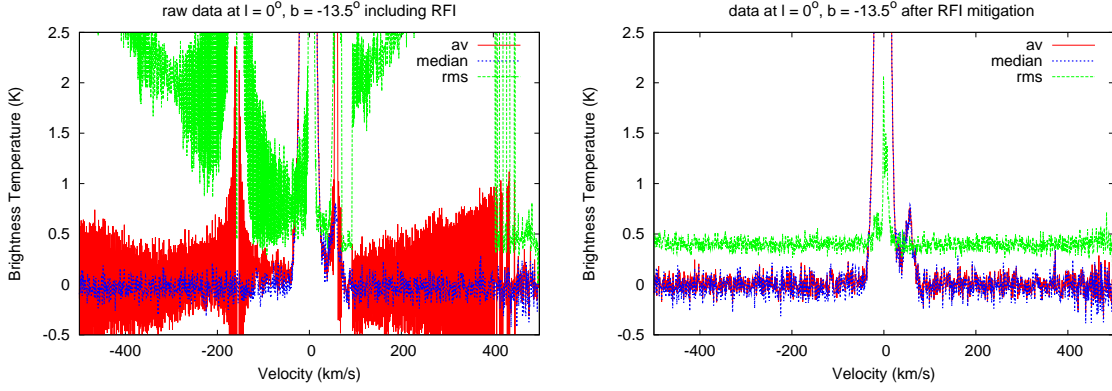
Figure 3: Data at the position $l = 0°$, $b = -13°5$ before and after RFI mitigation. The average is plotted in red, the rms scatter in green, and the median in blue.

3.3 RFI in emission line regions

RFI that falls in channels containing H I emission was usually removed successfully by the above procedure, except in regions with strong spectral gradients. Examples of such critical cases are positions with strong absorption lines. Sources with a continuum flux exceeding 200 mJy were excluded from RFI rejection.

To avoid any possible degradation of the signal due to an automatic RFI filtering, the median filters described previously were restricted to data with low brightness temperatures $T_{max} < 0.5$K. The remaining data were checked visually. Only a few cases showed some residual RFI. For these we increased the filter limits such that only channels with sufficient large rms deviation ΔT_{RMS} were affected. We verified that the remaining data were unaffected by the median filter. As a general rule we tried to avoid as far as possible any modification of the H I emission lines through the RFI filtering process. This implies, however, that some RFI may remain undetected and be present in the final data.

Some of the spectra were affected so severely by RFI that they needed to be removed completely. We eliminated all profiles whose mean rms within the baseline region exceeded the average noise level at its position by a factor of three. About 0.3% of the observed profiles were rejected for this reason.

4. Results

For HIPASS, as well as for the first data release of the GASS, imaging was performed by *Gridzilla* using weighted median estimation which is robust against RFI and other artifacts. For the second data release, we use separate and independent procedures for RFI rejection and imaging. As described in the previous section, our median filter is largely consistent with the HIPASS median filter algorithm for a radius $\varepsilon = 6'$. However, the major difference is that we apply this filter only to individual channels if two conditions apply: 1) an exceedingly large scatter is observed with a 3 times larger dispersion than the mean dispersion caused by thermal noise and 2) mean and median differ by more than one standard deviation. Only data that deviate from the median by $\Delta T_{RMS} > 2 \cdot \sigma(T_{RMS})$ are replaced by the median estimate.

A blind search for RFI at any given position demands quite elaborate calculations. In total 10^9 profiles need to be read in, some of them have to be updated after RFI excision. The original SDFITS database does not allow an efficient access to individual profiles. To optimize the I/O operations all data have been converted into a new database that allows fast random access. After this the cleaning process took several weeks.

All data have been inspected visually, only in very few cases the automated cleaning procedure failed. The algorithm is based on a robust determination of a clean median. This implies that less then 50% of the data need to be unaffected by RFI. Except for the time-independent RFI that was replaced by LAB data, this conditions was obviously met. The data were observed in two independent coverages with two different settings for the local oscillators. For such a setup time variable RFI would affect only 25% of the data, redundancy allows in this case to excise such data without degrading the date significantly.

We found that in total somewhat less than 0.5% of all data was affected by RFI and needed to be eliminated. This amount appears to be low but one should take into account that the 21 cm line of the Galactic neutral hydrogen is observed in the protected band. To demonstrate how dramatic the data are degraded by RFI we provide on our web server a movie that shows the RFI that was eliminated from the GASS database. Also data can be downloaded from http://www.astro.uni-bonn.de/hisurvey/gass/.

References

[1] Barnes, D. G., Staveley-Smith, L., de Blok, W. J. G., et al. *The Hi Parkes All Sky Survey: southern observations, calibration and robust imaging,* 2001, *MNRAS*, 322, 486

[2] Kalberla, P. M. W. et al., *The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections,* 2005, *A&A* 440, 775 [arXiv:astro-ph/0504140]

[3] Kalberla, P. M. W., Kerp, J. *The H I Distribution of the Milky Way*, 2009. *Annu. Rev. Astron. Astrophys.* 47, 27

[4] Kalberla, P. M. W. et al., *The Parkes Galactic All-Sky Survey. II. Stray Radiation Correction and Second Data Release,* 2010, *A&A* in press [arXiv:1007.0686]

[5] McClure-Griffiths, N. M., et al. *The Parkes Galactic All-Sky Survey. I. Survey Description, Goals, and Initial Data Release,* 2009, *APJS* 181, 398 [arXiv:0901.1159]

[6] Peirce, B. *Criterion for the rejection of doubtful observations,* 1852, *AJ*, 2, 161