# A NOVEL METHOD TO ESTIMATE THE RFI ENVIRONMENT

*David D. Chen[1], Chris Ruf[1]*

[1]University of Michigan, Ann Arbor, MI USA

## ABSTRACT

One of the greatest challenges in the evaluation of microwave radiometer RFI detection methods is a lack of suitable ground truth about the RFI environment; the false alarm rate can be readily estimated without it, but probability of detection cannot. A method is introduced to derive the probability distribution of RFI amplitude anywhere on the globe, thus characterizing the RFI environment. The method constructs low-level RFI characteristics which are difficult to detect from the high-level RFI characteristics that are more easily detected. The conceptual and theoretical basis of the method is presented, followed by possible future augmentations and improvements.

***Index Terms***— RFI, Aquarius, Radiometry

This paper introduces a method to estimate the RFI environment at any location on the globe, using the sea-surface salinity mission, Aquarius, as an example. Specifically, the RFI that the Aquarius radiometer is subjected to when its antenna is pointed at a particular location will be characterized. A systematic procedure can then be followed to derive any RFI statistic, which can be used to tune the RFI detection algorithm as a function of location, orbit node, polarization, etc.

Let the RFI contribution to the antenna temperature be represented by the random process $TA_{RFI}$. The process varies with geographic location but is assumed to be iid over time at any given location. This stationary process is fully characterized by a probability distribution function of the RFI signal amplitude.

## 1. MODELLING $TA_{RFI}$

For RFI levels near or below the NEDT level, reliable detection is very problematic. However, at higher levels, the RFI is typically straightforward to detect. Under the assumption of stationarity over time, but variability in space, low level, undetected RFI that enters the antenna pattern through its sidelobes at a particular location can be estimated by considering the appropriate neighboring locations through which it entered the mainbeam of the antenna at a much higher, and thus detectable, signal strength. Conversely, all detected RFI at the particular location can be assumed to have also entered the sidelobes in other locations at a lower signal strength. The ratio of detected (main beam entrant) signal strength to undetected (sidelobe entrant) signal strength would depend on the sidelobe level of the antenna pattern in the appropriate direction.

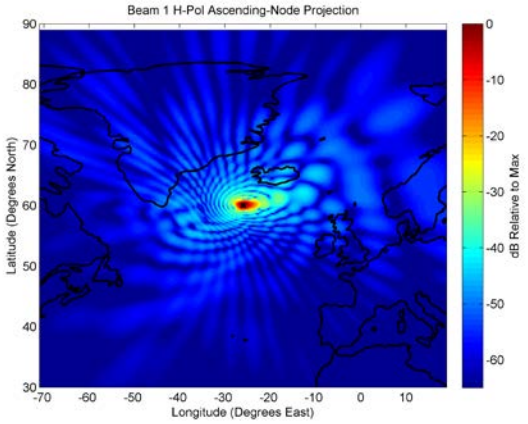


Figure 1: Antenna pattern projection with range effects, centered at 60N & 26W, which is denoted by X. $TA_{RFI}$ contains RFI contributions originating from many different locations, not just 60N & 26W.

This approach to estimating the low level RFI environment neglects the fact that viewing angles are slightly different when the RFI source is viewed through the mainlobe and through the sidelobe. Equivalently, we assume for now that all RFI sources radiate isotropically in some solid angle that includes all the directions under consideration. This assumption is later relaxed in Section 4.

Note that low-level RFI from the mainlobe cannot be included this way, but its effects are expected to be small – the RFI environments for the two orbit nodes shown by the peak hold maps in [1] suggest that much of the RFI enter through the sidelobes.

We zoom in to a particular location in Figure 1 to arrive at Figure 2, in which the location is denoted by a pentagram. Let us consider the region bounded by the outer circle in Figure 2, which corresponds to 96% of the beam fraction. This region is divided into $n$=25 patches, whose individual contributions can be summed to obtain $TA_{RFI}$:

$$TA_{RFI} = \sum_{i=1}^{n} TA_{RFIPatchi} \cdot wi \qquad (1)$$

where

$n$ is the number of patches,

$TA_{RFIPatchi}$ is the process modelling RFI originating from Patch $i$ only *when viewed through the mainlobe*, and

$wi$ is a weighting factor applied to $TA_{RFIPatchi}$ – it is treated presently.

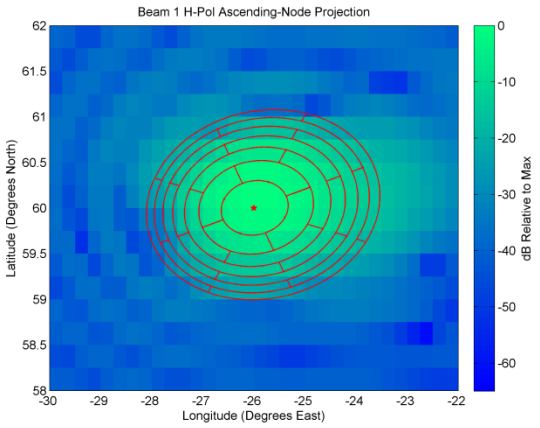


Figure 2: Region of consideration for deriving RFI distribution at a particular location.

Only one of the $n$ patches (the patch containing the pentagram) can be said to be viewed through the mainlobe, and all other patches are said to be viewed through the sidelobes. Except for the mainlobe patch, for which $wi = 1$, the $wi$'s for all other patches need to take into account the respective antenna gains and range.

RFI sources are too few to be taken collectively as an extended target; their probability distribution is far from Gaussian, as will be seen later. (In fact, had the distribution of been Gaussian, the RFI glitch detector would not be useful at all.) Thus, they must be treated as pointed sources, and the right formulation to use in the accounting of their power budget (and finding the $wi$'s) is the Friis transmission formula. Let us use $P_{rML}$, $G_{rML}$, and $R_{ML}$ to denote the received power, Aquarius antenna gain, and range, respectively, for the mainlobe patch, and use $P_{rSL}$, $G_{rSL}$, and $R_{SL}$ to denote the corresponding quantities in some sidelobe patch. Since power is proportional to antenna temperature in the microwave frequency regime, it can then be shown that the $wi$ for that patch is given by:

$$wi = \frac{P_{rSL}}{P_{rML}} = \frac{G_{rSL}}{G_{rSML}} \cdot \frac{R_{ML}^2}{R_{SL}^2}.$$

So for Patch $i$, $wi$ is the product of two factors: the antenna gain in the direction of Patch $i$ relative to the maximum antenna gain, and the ratio of mainlobe range to the Patch $i$ range, which we call the range factor. What was plotted in Figure 1 was precisely the product of these two factors, or $wi$.

Since a single $wi$ is used for the entire patch, it should be noted that the model becomes more accurate if each patch is smaller in area and the region is divided into a greater number of patches.

The following assumptions are made:

1. For every Patch i, $TA_{RFIPatchi}$ is an iid process.
2. For every Patch i and Patch j with i ≠ j, $TA_{RFIPatchi}$ and $TA_{RFIPatchj}$ are independent random processes.

Evidently, for every Patch i, $TA_{RFIPatchi}$ is completely characterized by a single distribution, which we will call the mainlobe distribution of Patch i, or $ML_i$ in short.

## 2. DETERMINING $ML_i$

To derive $ML_i$ (RFI characteristics at Patch i), we use only measured data with main lobe center in Patch i. We use the Aquarius detection algorithm for this task. The so-called clean mean [2] is the glitch detector's best estimate at the thermal emission signal. If a data sample is flagged, we subtract it the clean mean to find the detected RFI level; if it is not flagged, the detected RFI level is assigned a value of 0. If the result of the subtraction is negative, the detected RFI level is also taken to be 0. After this set is constructed, we compute the distribution of its elements, which we call $DT_i(x)$. (References to "distribution" in this work should be interpreted as the probability mass function.) x is the an element of the domain of the distribution, which is the set of all possible RFI levels. $DT_i(x)$ is derived for both orbit nodes. The $DT_i(x)$ for the patch containing 59N 26W, which we call DT at 59N 26W, is shown in Figure 3 for the ascending node. Note DT(0)=97.7%, which means RFI (of any level) occurs only 2.3% of the time.

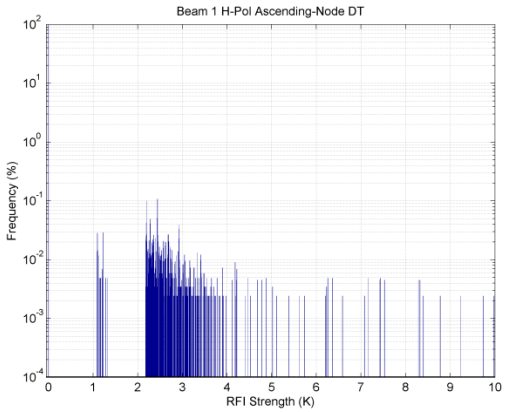


Figure 3: Ascending-node DT at 59N 26W. Ascending and descending orbit nodes are considered separately due to the RFI environment being different for both [1]. Note the log scale on the y axis, and the high value at 0 (97.7%), which is hard to see. The cutoff seen around 2.18 K is due to the TD threshold in the Aquarius RFI algorithm. Only data samples deviating from the clean mean by more than this threshold are flagged.

As stated earlier, $ML_i$ for Patch i is defined as the distribution of RFI sources residing only in Patch i. For any location, the viewing angles for ascending and descending orbits are different in general. But if a source (in Patch i) is isotropic, it will contribute equally to both the ascending and descending distributions. Thus, for each nonzero x, $ML_i(x)$ is taken as the minimum of the ascending $DT_i(x)$ and descending $DT_i(x)$. $ML_i(0)$ is assigned a value so that

$ML_i(x)$ is a valid distribution. (In practice, because RFI levels may shift slightly, a window of 0.5 K is used to compare the ascending and descending $DT_i$'s.)

This simple algorithm is not perfect. There are some cases where a source is counted in $ML_i$ even if it does not reside in the main beam patch (if there are contributions from sidelobes in both ascending and descending orbits, for example). Also, some directional sources in the mainlobe can be missed. Later in Section 4, we discuss some of these caveats and show how they can be taken into account.

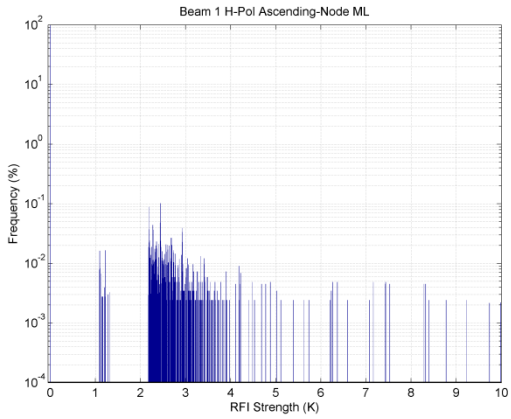


Figure 4: Ascending-node ML at 59N 26W.

## 3. ADJUSTING FOR AREA SIZE

In Equation ( 1 ), $TA_{RFI}$ is modeled as the sum of $n$ RFI patch processes, with each patch having an arbitrary but well-defined area. $TA_{RFIPatchi}$ is dependent on the size of area it characterizes. Besides revealing the magnitude of the RFI sources in the patch, it conveys the frequency at which sources emit, or how densely populated the RFI sources are in the patch. If the sources in the patch emit more frequently, or if there are more sources in the area, the value of the distribution at 0 would decrease, and the distribution would increase at some non-zero value(s). The distribution would change in the same way if it represented a bigger area because there would be more RFI sources in the area the distribution characterizes. We determine this effective area corresponding to the $DT_i$ and $ML_i$ derived thus far as follows.

In the histogram of $ML_i$ shown in Figure 4 above, we can differentiate between two general regions in the RFI level. From about 4 K and higher, the RFI strengths are sparsely distributed and the probability is relatively constant. These are the high-level RFI sources. Below 4 K, the lower-level RFI strengths are much more continuous with small and few gaps. There is also a gradual increase in the probability of occurrence as the strength decreases. If we assume the lower-level RFI are roughly constant in strengths, one cause of the gradual increase in probability is the weighting by the antenna gain: in the mainlobe, there is more area which is seen through a smaller gain than a higher gain – for example, there is only a single point at which the antenna gain reaches its maximum. Let's consider the point at the center of the mainlobe which detects RFI with the highest antenna gain. It is through this point that the highest value of lower-level RFI will be observed. This is seen to correspond to about 4K. Since the cutoff due to the TD threshold occurs at 2.18 K, the antenna gain ratio at the cutoff is approximately $2.18\,\mathrm{K} / 4\,\mathrm{K} \doteq 0.55$. Using the antenna pattern data, this is seen to correspond to a zenith/polar angle of about 2.9 degrees. The effective area can then be readily computed from the projection of the antenna pattern on the ground. The effective area is dependent on the TD threshold used earlier to derive the DT distribution in Figure 3.

Once the effective area of our $ML_i$ is determined, we may wish to know the RFI histogram of an area smaller than the effective area, because as mentioned in Section 1, the more finely we divide the region of observation, the more accurate the composite histogram of $TA_{RFI}$ will be. One method of doing this is suggested in Section 5. Here we introduce a way to approximate the distribution of a smaller area.

Let Patch i be divided into 2 sub-patches j and k with equal areas. Thus, $TA_{RFIPatchi}$ is the sum of two random variables, $TA_{RFIj}$ and $TA_{RFIk}$, which describe the mainlobe RFI distribution from sub-patches j and k, respectively: $TA_{RFIPatchi} = TA_{RFIj} + TA_{RFIk}$. We assume $TA_{RFIj}$ and $TA_{RFIk}$ to be iid. Thus, they are characterized by the same distribution $ML_j = ML_k$. As mentioned earlier, RFI occurs very infrequently (this is true for other locations over the ocean as well). Using this fact, and with $ML_i$ is known, it can then be shown that the following is a good approximation for $ML_k$: For each nonzero x, $ML_k(x) = ML_i(x)/2$. $ML_k(0)$ is assigned a value so that $ML_k(x)$ is a valid distribution.

This is what we would expect intuitively: if a sub-patch is ½ of the area of the original patch with known RFI distribution, any non-zero RFI level will observed approximately half as frequently. At the same time, the probability of no RFI emission increases to ensure the probabilities in the histogram sum to 1. The approximations become worse when the number of divisions or sub-patches is increased.

This approximation is generalized to derive the $ML_l$ of a sub-patch l that is has an area of f*effective area of Patch i. For each nonzero x, $ML_l(x) = ML_i(x)*f$. $ML_l(0)$ is assigned a value so that $ML_l(x)$ is a valid distribution. In this work, f is kept between 0.5 and 1.

## 4. THE FINAL COMPOSITE DISTRIBUTION

$ML_i$ thus derived is our best estimate of the distribution of $TA_{RFIPatchi}$. As Equation ( 1 ) indicates, the $TA_{RFI}$ is a weighted sum of all the $TA_{RFIPatchi}$'s, so its distribution can be computed by convolving the weighted $ML_i$'s. Let us call the distribution of $TA_{RFI}$ thus derived

ISO. In Section 2, we estimated the mainlobe RFI, $ML_i$, by taking the common portion of ascending node $DT_i$ and descending node $DT_i$. This method assumed that RFI sources are isotropic. Some examples of RFI sources that can be missed in $ML_i$ and subsequently in ISO are:

1. Directional sources in the mainlobe and sidelobe patches, and
2. Non-stationary sources that exhibit some time correlation with the 7-day repeat orbit of Aquarius and have different characteristics in the two orbit nodes.

Fortunately, the RFI in these cases are detected and included in DT of the mainlobe patch for the respective orbit node. Thus, to take into account these missed sources, we improve ISO to create the final, composite distribution known as CP as follows. For each nonzero x, CP(x) is the maximum of DT(x) and ISO(x). CP(0) is assigned a value so that CP(x) is a valid distribution.

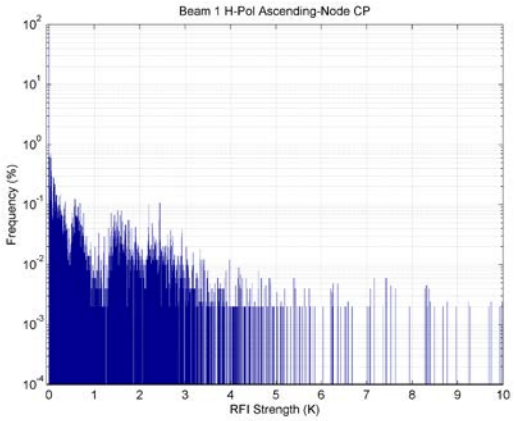


Figure 5: Ascending-node CP for 59N 26W

Finally, it should be noted that the power of an isotropic RFI point source, due to different ranges to ground, will appear differently to the 3 beams of Aquarius. Beam 3, with the largest off-nadir angle, will report a lower power for this source than Beams 2 and 1. This is not a result of the lack of calibration between the 3 beams. If a piece of homogeneous natural terrain is large relative to the size of the footprint, and has a TB of 15K (without RFI), the 3 Beams, which have been calibrated, will each report a TA of 15K . But this does not apply to RFI, which are point sources. So this "inter-calibration" of point sources, taking into account the effects of range and effective area, was done in the generation of $DT_i$ in Figure 3.

This completes the characterization of $TA_{RFI}$. A summary of entire derivation process is shown in Figure 6.

## 5. FURTHER WORK

Introducing more patches, and over a larger region than shown in Figure 2, would improve the accuracy of the method. We introduced a method in Section 3 to estimate the RFI distribution of a sub-patch with an area smaller than the effective area. However, the assumptions used to derive this approximation break down if the effective area is divided into too many sub-patches.

For division into n iid sub-patches, $TA_{RFIPatchi} = \sum_{i=1}^{n} X_i$, where $X_i$ is the random variable characterizing the $i$th sub-patch. The goal is to solve for the distribution of $X_i$ given the distribution of $TA_{RFIPatchi}$. One solution might be to make use of the Fourier Transform (or characteristic function of the random variable), with the constraint that each distribution must take on non-negative values for all elements of its domain. Existence of solutions is not guaranteed. For example, if $TA_{RFIPatchi}$ has a distribution with support only at 0K, 3K, and 5K, it can be shown no solution exists for $n$ =2. So for some distributions of $TA_{RFIPatchi}$, the assumption of iid $X_i$'s may not be valid. Or perhaps the distribution of $TA_{RFIPatchi}$ is missing some values, due the finite amount of data used to derive the distribution. A method of finding the "best" solution can be investigated in future work.

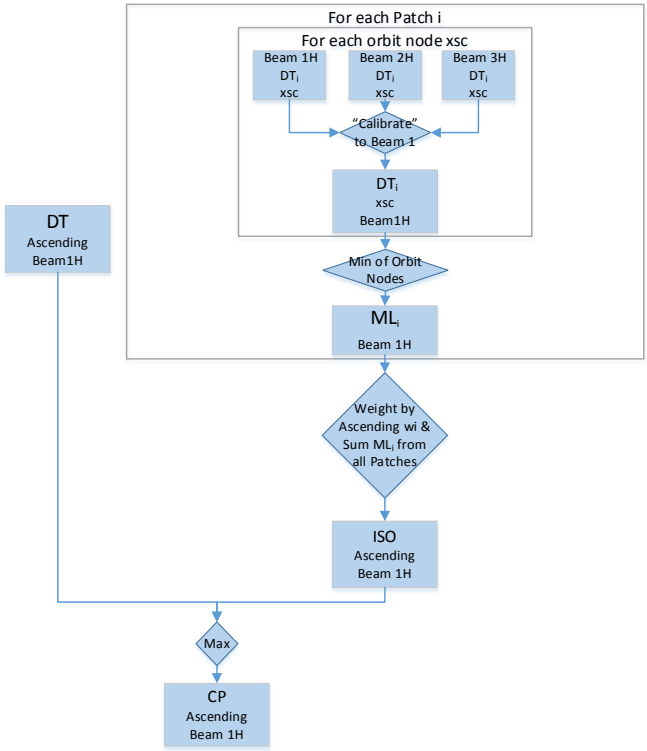


Figure 6: Derivation of the distribution of $TA_{RFI}$ at a particular location for Beam 1, H-pol, ascending node

## 6. REFERENCES

[1] Ruf, C., D. Chen, D. Le Vine, P. de Matthaeis and J Piepmeier, "Aquarius Radiometer RFI Detection, Mitigation and Impact Assessment," Proc. 2012 IGARSS, Munich, GERMANY, pp. 3312-3315, 23-27 July 2012.

[2] S. Misra and C. S. Ruf, "Detection of Radio-Frequency Interference for the Aquarius Radiometer," Trans. on Geosci. Remote Sens., 46, 3123-3128, 2008.