Correlations of spectral intensity fluctuations: Application to radio frequency interference mitigation

Avinash A. Deshpande

Arecibo Observatory, National Astronomy and Ionosphere Center, Arecibo, Puerto Rico

Raman Research Institute, Bangalore, India

[1]   This paper describes a technique based on correlating fluctuations in spectral intensities. When applied to intensity data recorded as a function of time and frequency, the result can be viewed in the form of a correlation map, wherein the temporal correlation between fluctuations in every possible pair of spectral channels is represented. In addition to serving as a useful diagnostic tool for the measuring system itself, such a correlation map offers a wealth of information on spectral channels affected by radio frequency interference (RFI) and intermodulation products, if any. Although such estimations are inherently less sensitive than those using voltage correlations, the technique promises much broader applicability since intensity data are more commonly available. The strengths and weaknesses of this technique with respect to RFI mitigation are discussed and are illustrated with real examples. An extension of this analysis to polarization data is also explored. A promising way to isolate RFI on the basis of its highly polarized nature is proposed, and relevant practical issues are discussed. The present study suggests that high-resolution time-frequency data for the full Stokes parameters allow effective excision of RFI (by typically more than 20 dB), particularly for observations of unpolarized astronomical sources. This method is exceedingly effective in situations where a large fraction of the data is affected by polarized RFI, where a robust statistics approach fails.

1. Introduction

[2]   Correlation of signal voltages as a means of detecting and estimating common mode signals is well known in radio astronomy and other fields. The most routine use of the correlation operation, as in interferometry or spectral estimation, involves signal voltages as inputs, rather than signal intensities which in fact are the results of a voltage correlation. For many years now, it has been possible to record intensity data with nominal online integration, and any ''matched filtering'' is then attempted as a part of subsequent off-line analysis. This provides considerably more flexibility and sophistication to the signal processing as long as the relevant data rates are not prohibitively large. It also allows detailed editing of data, an unavoidable step when data are affected by man-made interference. Although reference antennas and voltage correlations are beginning to be increasingly used to help assess and excise radio frequency interference (RFI), we still need tools that are applicable to conventional intensity data which are more routinely available.

[3]   In this short paper, we focus on this type of data, that is, recorded intensities (signal power) as a function of time and frequency, and examine certain aspects of possible correlation between fluctuations in the intensities and their relevance to RFI detection and excision.

[4]   In section 2, we introduce a technique that allows examination of sets of mutual correlations between intensity fluctuations in each available spectral channel. We discuss how such a correlation set offers a sensitive diagnostic tool for the measuring system itself and a

wealth of information on spectral channels affected by RFI and any intermodulation products. A few illustrative examples of resultant correlation maps are presented. The various signatures which provide useful insight into the nature of possible nonastronomical signal contributions are discussed.

[5] Section 3 proposes a technique for interference excision that is based on the expected contrast in the polarization characteristics of man-made interference with that of most commonly encountered astronomical signals. We also present some results of tests, and identify the regime in which our technique may be most effective. One recipe for interference excision is outlined in section 4. Section 5 summarizes the new approaches and discusses their implications.

2. Correlation of Fluctuations in Intensity

[6] In this section, we introduce a correlation technique which allows us to examine the correlation between temporal fluctuations in each spectral channel with itself and with those in all other available channel. This technique, in the context of pulsar intensity correlations across rotational longitudes, was first suggested by *Popov and Sieber* [1990]. For the data in each channel, we estimate a robust mean and RMS across the relevant time sequence of its data. By "robust" we imply that outliers, if any, are excluded (as far as possible) from the computation. The "robust" mean estimate generally agrees well with the corresponding 'median' estimation while still benefiting from the desirable averaging effect (particularly important when the sample size is small). From each of a pair of time sequences, the respective robust mean is subtracted and the resultant sequence normalized with respect to the robust RMS. An average correlation for that pair is then simply the average product of the two sequences. In the absence of any outliers, we expect such a correlation to be close to (and on the average equal to) unity when the sequences are copies of each other. However, any outliers would be apparent (at least) along the diagonal, with the corresponding (auto)correlations exceeding unity. Following this procedure, a map of correlations can be obtained wherein an $N_s \times N_s$ matrix of correlations (resulting from N_s spectral channels) can be examined together. The diagonal elements of such a map correspond to autocorrelations or cross correlations (depending on whether a given set of time-frequency data is correlated with itself or with another set, say, for a different parameter), while the off-diagonal elements, always correspond to cross correlations. It is worth noting that, in the absence of correlated outliers and temporal variations in system temperature and/or gain, the cross correlations are expected to be ~ 0, on the average. The subtraction of the mean and normalization

by the RMS described above makes the correlation measures less sensitive to gain variations as a function of frequency.

[7] As an example, Figure 1 shows a correlation map obtained using a short sequence of Stokes-I L band spectral data (20 sec span, 100 MHz bandwidth centered at 1270 MHz, 2048 channels, three-level sampling, 10 ms dumps) from the Arecibo Telescope. The middle plot shows the matrix of correlations, while the left and the bottom plots show the average spectra for the respective sets, which for this example are identical. Here, most of the diagonal elements can be seen to have (auto)correlations close to unity, with most of the other elements having (cross) correlations close to zero, as expected. However, there are significant deviations at the locations of certain spectral channels. These can be readily seen along the diagonal and identified as due to certain well-known RFI sources at Arecibo. For example, the strong narrow features at about 1241 and 1256 MHz are due to the Punta Salinas radar. The Aerostat radar in Lajas also shares these frequencies, but blanks itself when pointing toward the Arecibo Telescope. The narrow feature at about 1290 MHz is due to the Ramey radar. However, the correlation map shows much more structure associated with these and a few other frequencies. The pairs of features at 1232 and 1247 MHz (and that at 1286 and 1310 MHz) are most likely to be associated with the Punta Salinas radar, judging by their apparent correlation with the stronger features. Note that the latter pairs do not show themselves clearly in the average spectra. However, their dominant presence in the correlation map makes it easy to identify them and to assess the possible common origin that they appear to share within the pair as well as with the stronger features at 1241 and 1256 MHz. The fluctuations at these pairs of frequencies show correlation across almost the entire band (apparent along a cross in the map), although at a much lower level. Such a signature would be expected if the measuring/receiver system suffers gain compression in the presence of the time varying RFI. Apparent gain compression is a manifestation of system's inability to output correspondingly more power when the power at its input increases (for example, because of a strong RFI). Since the normalized spectral response of the system is normally unaffected by gain compression, any power level increase in the RFI channels thus causes a finite drop in the output power for the rest of the band, preserving the total output power. The cross correlation along the cross is expected then to be on the average negative, implying anticorrelation. The negative correlations (blue bands) seen at the edges of the map have a similar origin, however their apparent enhancement is an artefact of the falling gain at the edges. It is interesting that the RFI pair with a weak average contribution appears to show more serious gain compression. We interpret this as the
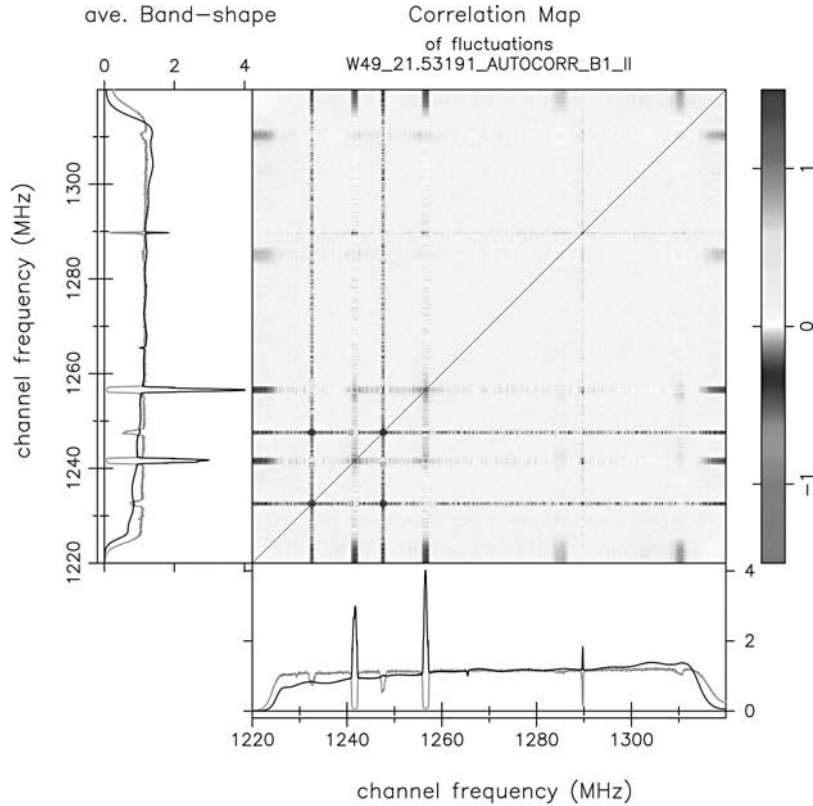
Figure 1.   Correlation map of fluctuations in spectral intensity of Stokes I. See text for details. See color version of this figure at back of this issue.

particular RFI having a larger peak intensity but a very much shorter duty cycle in comparison with the apparent stronger features, resulting in a relatively small net contribution to the average.

[8] Possible variations in the overall gain of the system or in the system temperature would contribute a positive correlation pedestal across the entire map. No significant pedestal is apparent in the above map, indicating the absence of any significant gain or system temperature variation during the short duration of the sequence. Although (fortunately) absent in the present example, the presence of intermodulation products resulting from such strong RFI pairs and their interrelationships would also be apparent in correlation maps as off-diagonal contributions.
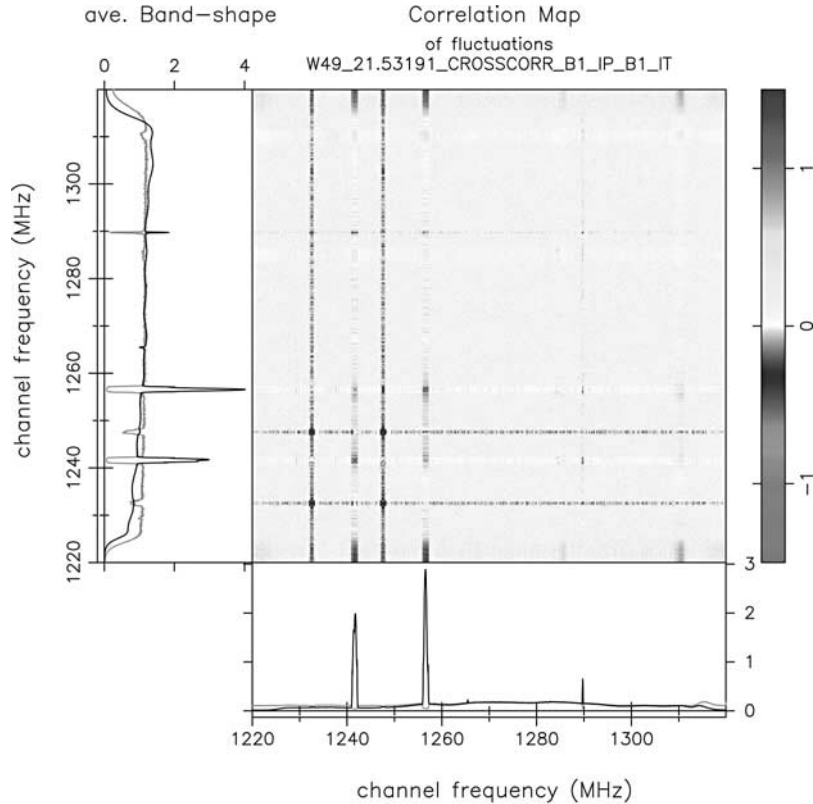
[9] The color line plots in the bottom and left plots show what could be termed to be the "efficiency" of the measuring system. The plotted quantity is simply the ratio of the observed signal-to-noise ratio (i.e., mean/RMS) to that expected ideally. Except at the locations of the RFI (strong or weak) and at the band edges, the efficiency is at about the level expected for the three-level digitization. This simple measure, which is just

another version of the Allen variance, is easy to compute and is very useful for identifying instabilities in general, and RFI in particular.

[10] The accuracy of correlation estimation naturally depends on the number of independent sample pairs (N_t) used, where the RMS uncertainties in the correlations are proportional to $1/\sqrt{N_t}$. Since the sampling intervals at which intensities are recorded are usually many orders of magnitude longer than the inverse of spectral resolution, the number of independent samples available per unit time is normally rather small. This makes the intensity correlation estimation inherently less sensitive than voltage correlation. However, our technique promises much broader applicability since intensity data are more commonly available.

3. Polarized Nature of RFI: Promising Criterion for RFI Excision

[11] The above procedure can be readily extended to obtain other possible correlation maps where the data on any pair of the two polarization or Stokes channels can

Figure 2. Cross-correlation map of fluctuations in polarized and total intensities. The corresponding average spectra are shown in the bottom and left plots. See color version of this figure at back of this issue.

be examined. Such maps made for the data used in Figure 1 reveal that the RFI is highly polarized. To illustrate this, in Figure 2 we present the cross-correlation map between the polarized and the total intensities. As far as the interference features are concerned, these cross correlations follow closely the pattern seen in Figure 1. This is not at all surprising since most communication signals are intrinsically 100% polarized. Similar maps for other Stokes pairs (not shown here) are consistent with this observation.

[12] This opens up an attractive possibility for isolating the RFI contribution from that of an unpolarized astronomical signal. To quickly assess this possibility we examine our nominal estimate of the unpolarized intensity, computed simply as $I - P$, where I is the total intensity (Stokes' I) and $P = \sqrt{Q^2 + U^2 + V^2}$ gives estimate of the polarized intensity.

[13] Figure 3 shows the resultant dynamic spectrum for the nominal estimate of the unpolarized intensity and the corresponding time and frequency averages. Note the reduction (by about 20 dB) in the contribution of RFI to

the average spectrum. The "absorption"-like feature (associated with an OMT resonance of the L band feed) is now more apparent. The associated correlation map (not shown here) continues to have a similar appearance as it retains its sensitivity to the residual correlation. However, a significant improvement in the "efficiency" at the RFI channels is evident. The improvement seen here (in Figure 3) is only partial, although remarkable considering that the data has not been calibrated for instrumental gain/polarization response. We show three sample time sequences (each corresponding to a narrow frequency channel) in Figure 4.

[14] It is easy to see that considerable improvement in RFI isolation should be expected consequent to proper polarization channel calibration. Note that the so-called instrumental polarization, if any, would result in underestimation of the unpolarized contribution, particularly in the RFI-free channels. However, in general, in the presence of RFI or any other polarized signal, the instrumental polarization would result in depolarization and the consequent leakage of polarized intensity into the
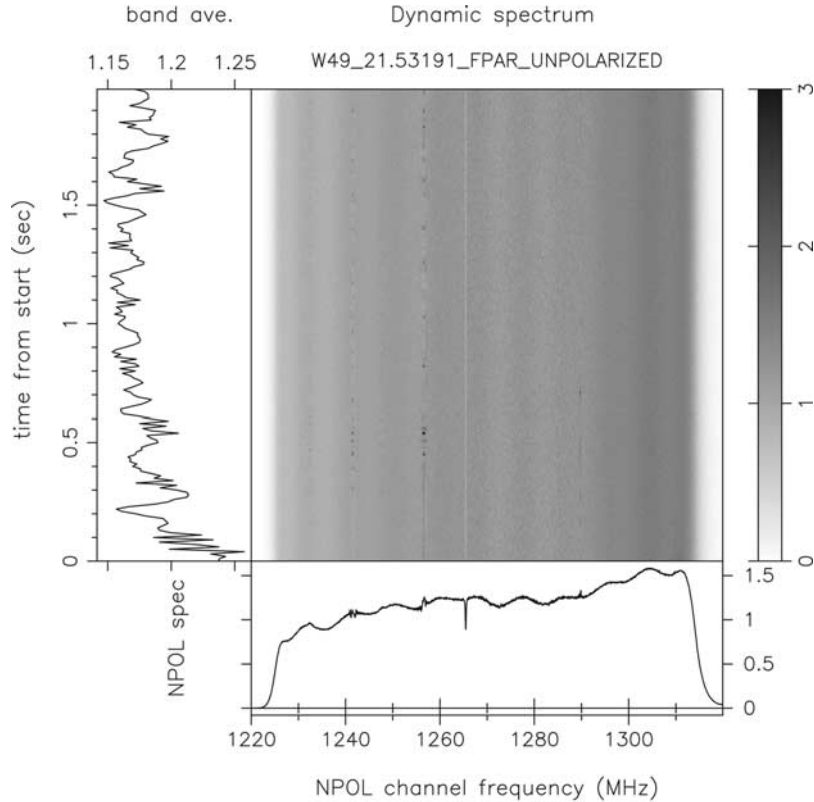
Figure 3. Dynamic spectrum of unpolarized intensity. The associated average spectrum and time sequence are displayed in the left and bottom plots, respectively. Note the remarkable reduction in the interference contribution. See text for more discussion.

apparent unpolarized component. We have tried to assess this with a very preliminary gain calibration performed on this data, and our results confirm this expectation. It is worth noting that any calibration details that amount to a mere rotation of the polarized Stokes vector (i.e., [Q, U, V]) are not important in the present context. However, calibration of the relative gains of the native polarization channels of the receiver and any other aspects that modify the magnitude of the Stokes vector are important. Even in the case of appropriately calibrated data, one is still likely to face the following difficulties.

[15] 1. Local scattering off the telescope structure (or other possible reflections including ground reflections, etc., during propagation) may lead to mild but detectable depolarization of an otherwise fully polarized signal. However, such depolarization can occur only if a signal has finite ''noise bandwidth'' and if the associated relative delays are significant with respect to the inverse of the signal bandwidth. Therefore, for man-made deterministic signals, such scattering may modify the state of polarization of the signal, but it will remain 100% polarized. However, in practice, the contribution from man-made signals is superposed on those from the receiver and sky noise which may have finite degree of polarization owing to local reflections or is intrinsic to the source.

[16] 2. Depolarization can also occur if, for example, more than one source of RFI happens to be present, or if the polarization state of a given RFI source varies within one or more of the time-frequency bins. This suggests a need for sufficiently high time and frequency resolution.

[17] 3. Computation of quantities such as the polarized intensity (P) involves a ''squaring'' operation, and that leads to a statistical bias primarily from any noise associated with the unpolarized intensity (as it implicitly contributes to uncertainty in estimation of the polarized intensity). It is possible to correct for such a bias [e.g., *Wardle and Kronberg*, 1974], in the context of the Ricean bias), but only in a statistical sense. Relevant analytical expressions for estimating uncertainties in the polarization parameters in the synthesis imaging and single-dish cases are available [e.g., *Deshpande*, 1994]. The uncertainties in the presence of a polarized contribution are shown to be larger than otherwise, although
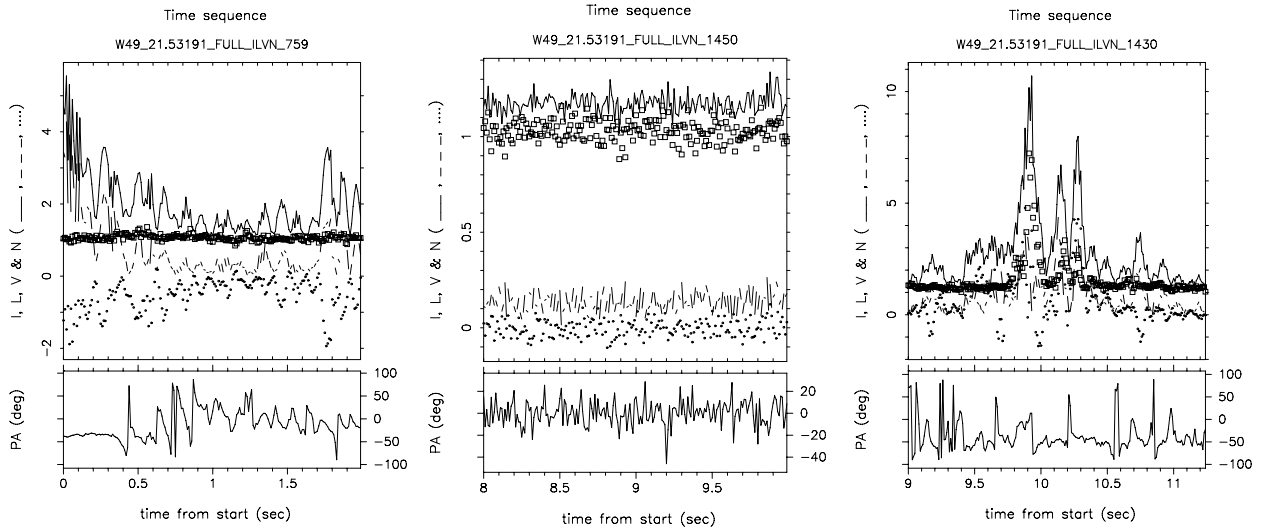
Figure 4. Three examples of time sequences of the total intensity I (solid line) and the estimated unpolarized intensity (open squares). The linearly (dashed line) and circularly (dots) polarized components are also indicated for comparison. The plot on the left illustrates the case when removal of a polarized contribution has amounted to removing the RFI contribution, while the case on the right shows that the RFI contributions may sometimes have an apparent unpolarized component. The middle plot shows a control sample corresponding to a spectral channel which, we believe, is unaffected by RFI. The average difference between the total and the unpolarized intensities is due to statistical bias in the estimation of the polarized intensity.

the increase is not linear with increasing degree of polarization. The magnitude of such corrections can be relatively large when the time-bandwidth product associated with the data samples is not large. Unfortunately, the requirement for a large time-bandwidth product is in potential conflict with the requirement for finer time and frequency resolution. In any case, it is important to remind ourselves that when the time-bandwidth product equals unity, the apparent polarization is 100%. It is only through the superposition of the polarization states associated with many independent samples that the unpolarized intensity, if any, starts to become apparent. Another way of stating this would be that even if a signal is fully unpolarized, the magnitude of the statistical bias in the apparent polarized intensity will match the total intensity when the time-bandwidth product is unity.

[18] Overall, the prospects for using the unpolarized intensity as an "RFI-free" quantity are encouraging, and the estimation procedures need to take into account the above and other relevant aspects.

4. Possible Recipe for Interference Excision

[19] It is clear from the above discussion that full-Stokes time-frequency data, with suitably high spectral and temporal resolution, are desirable. Given such a data

set, the various correlation maps (as well as the measures of system "efficiency") can be computed and examined to identify spectral regions affected by interference. These sensitive measures allow one to determine whether a given spectral feature is astronomical or not. Note that a stable spectral line feature of astronomical origin, regardless of its intensity, would not show up in the correlation map, or even in the "efficiency" spectrum, since the associated temporal fluctuations would follow the expected statistics for a sky contribution. Thus the correlation analysis offers a sensitive filter to distinguish between the statistics associated with astronomical signals and those resulting from man-made RFI.

[20] If the RFI only rarely appears in a particular set of spectral channels, then a commonly used simple robust mean computation of intensities from each of the temporal sequences for each individual channel separately would give an RFI-free estimate of spectral intensity across the spectrum. However any persistent RFI, such as is seen in the examples shown in the left and right plots of Figure 4, does not allow meaningful estimation of a robust mean that would relate to an RFI-free level. In such cases, estimation of the unpolarized intensity offers a remarkable advantage. It would also seem to have advantages when the data represents that taken with the telescope scanning, that is, in situations where a robust

mean is not relevant. Before estimating this quantity, suitable channel gain calibration (across the spectrum) needs to be performed. To minimize the effects of any statistical bias that otherwise results in an underestimation of the unpolarized component, a suitably scaled and offset-corrected version of the estimated time-sequence of the polarized intensity can be subtracted from the total intensity sequence of the corresponding spectral channel. The relevant scale factor would be less than or equal to unity, and can be estimated from the observed cross-correlation of the polarized intensity sequence with the total intensity sequence for each channel separately. This works well when the polarized intensity (e.g., RFI), in any given channel dominates the unpolarized component. However, in general, it is possible to estimate and correct for the statistical bias by using the theoretical estimate of the associated variance at each of the sampled points [e.g., see *Deshpande*, 1994].

[21] Even if the RFI contribution does not get removed entirely from the time sequence of the unpolarized intensity, such time sequences will be free from "persistent" RFI, and a simple robust mean estimation for each of the spectral channels is then able to excise the residual RFI contribution from the "average" spectrum. This can be appreciated from the situation encountered in the example shown in the right plot of Figure 4.

5. Summary

[22] We have presented, via a few real examples, a technique based on correlations of fluctuations in spectral intensities for highlighting and identifying weak RFI features and their possible common origins.

[23] Through studies of such correlations, a remarkable commonality between temporal fluctuations in polarized and total intensities is identified. On the basis of this and other related tests, we show that an effective way of removing RFI is to estimate the unpolarized intensity. This simple technique will be very useful in Galactic and extra-galactic HI and recombination line observations, as well as for the continuum mapping of unpolarized radio sources. For continuum observations of polarized sources (including pulsars) it is also possible to benefit from

these techniques, wherein the RFI affected channels are identified by employing the correlated intensity approach and/or through the deterministic polarization signature, and excluded when estimating the relevant band-averaged quantities. One can also appeal to the modulation in the spectrum due to Faraday rotation (i.e., finite rotation measures) to separate any polarized sky contribution from RFI. In general, recognizing RFI as a highly polarized contaminant provides a remarkably effective criterion for estimating and removing its contributions, even in situations of persistent RFI where robust statistics approach fails. However, for effective application of the RFI-excision techniques discussed here, it is essential that high-resolution data in both time and frequency be recorded for all the Stokes parameters.

References

Deshpande, A. A. (1994), Source-noise in radio synthesis images of polarized sources, in *Very High Angular Resolution Imaging: Proceedings of the 158th Symposium of the International Astronomical Union*, edited by J. G. Robertson and W. J. Tango, pp. 95–99, Springer, New York.

Popov, M. V., and W. Sieber (1990), Correlation of pulsar radiation intensity fluctuations at various longitudes of the mean profile—Analysis of observational data, *Sov. Astron.*, *34*, 382–409.

Wardle, J. F. C., and P. P. Kronberg (1974), The linear polarization of quasi-stellar radio sources at 3.71 and 11.1 centimeters, *Astrophys. J.*, *194*, 249–255.
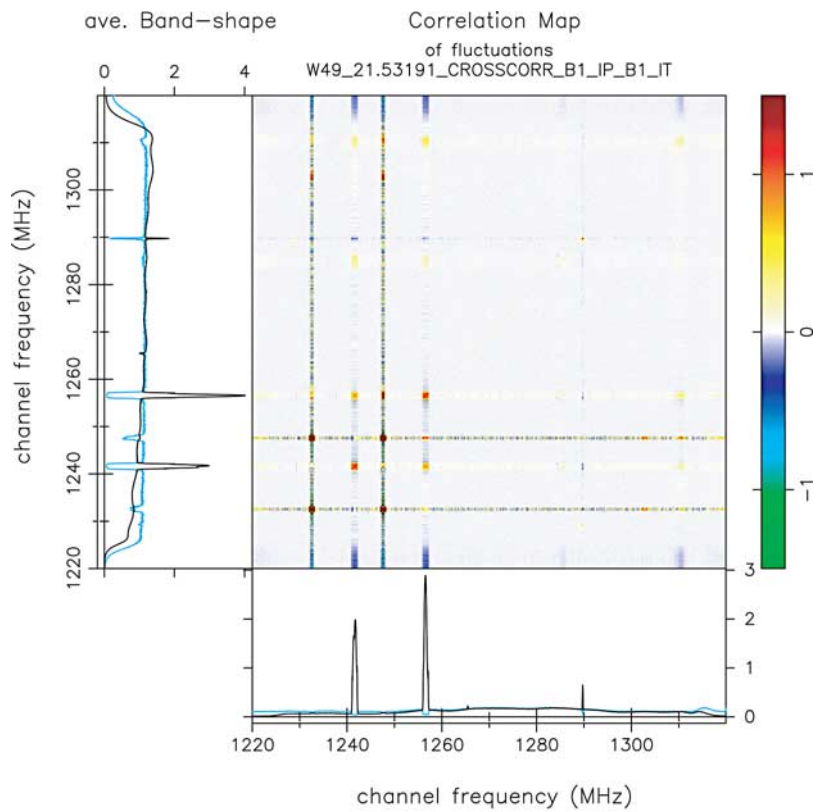
A. A. Deshpande, Arecibo Observatory, NAIC, HC 3 Box 53995, Arecibo, PR 00612. (desh@naic.edu)
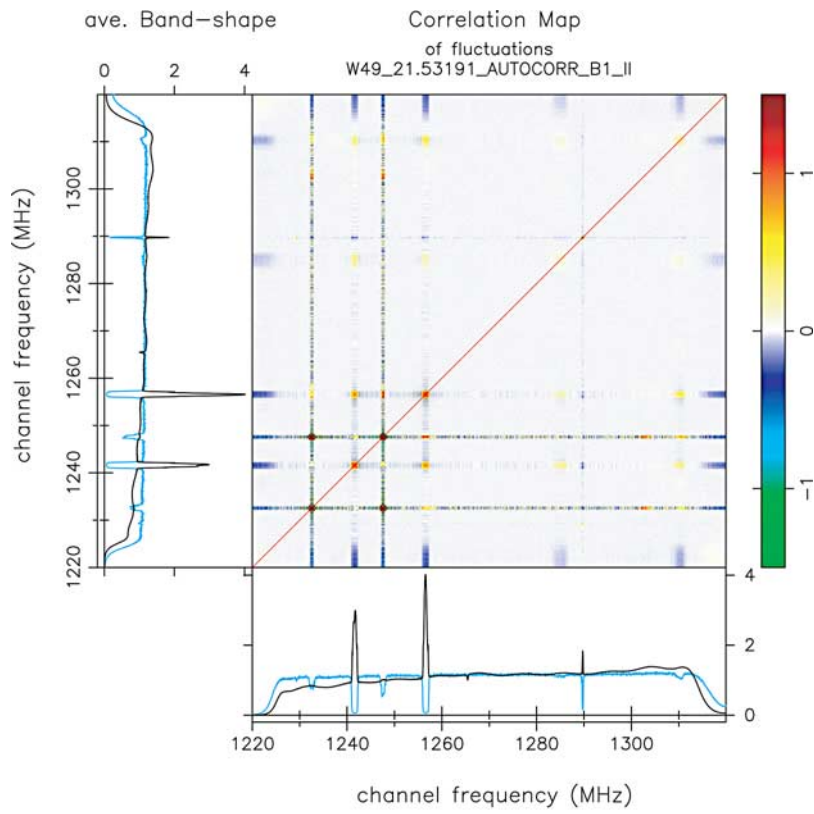
Figure 1. Correlation map of fluctuations in spectral intensity of Stokes I. See text for details.

Figure 2. Cross-correlation map of fluctuations in polarized and total intensities. The corresponding average spectra are shown in the bottom and left plots.