N87-10520

101

8.5.1 INTERFERENCE DETECTION AND CORRECTION APPLIED TO INCOHERENT-SCATTER RADAR POWER SPECTRUM MEASUREMENT

18785

W. P. Ying*, J. D. Mathews, and P. K. Rastogi

Electrical Engineering and Applied Physics
Case Western Reserve University
Cleveland, Ohio 44106

ABSTRACT

A median filter based interference detection and correction technique is evaluated and the method applied to the Arecibo incoherent-scatter radar D-region ionospheric power spectrum is discussed. The method can be extended to other kinds of data when the statistics involved in the process are still valid.

INTRODUCTION

The expression for the D-region ionosphere incoherent-scatter radar (ISR) power spectrum has been a well-known quantity (DOUGHERTY and FARLEY, 1963; TANENBAUM, 1968; MATHEWS, 1978, 1984a), from which more physical parameters can be inferred (MATHEWS, 1984b) than if the power profile measurement is carried out alone. However, a real-time power spectrum measurement was not possible until the computer power advanced so that an efficient fast Fourier transform can be realized into the array processor. Tepley (TEPLEY et al., 1981) first reported a successful experiment of D-region power spectrum using the 430-MHz ISR at Arecibo. Although the collision-dominated power spectrum shape can be seen easily from the raw data, interference can be a serious problem which sometimes even overwhelms the spectrum totally.

Several interference removal techniques have been devised, among them the commonly used method to model a theoretical data set, and then divide the experimental data with the theoretical data to obtain a 'flat', 'noise-like' data sequence which then allows easier detection and removal of outliers from the data. Another useful method is to form two complementary data sets by summing and subtracting the experimental data with the theoretical data, and then sum two complementary data sets to get rid of the outliers (RASTOGI, private communication). In both of these approaches, the performance depends totally on how accurately the theoretical data resembles the true data, i.e., a good *a priori* knowledge of the real data is required, but this condition is seldom met.

The method we bring out here, which is named the Template Process, is based on the concept of median filtering (RASTOGI, 1983). Even though we applied this method to the D-region power spectrum, it is independent of the shape of the data and no prior knowledge is necessary, therefore is suitable to process other kinds of power spectra as long as the statistical properties of the data can be described by the assumptions that we made to form this method. Consequently, it is a good practice to treat the following discussion as a generalized idea and we use the D-region power spectrum only as an example.

TEMPLATE PROCESS -- METHODOLOGY

Median Filtering. A time integrating method (or time averaging) can enhance the estimate of the return signal by increasing the signal-to-noise ratio. But this applies only to the additive Gaussian noise channel. If other

---

*Presently at Arecibo Observatory, Arecibo, Puerto Rico 00613 signals, like interference, are involved in the integrating process, then the estimate is biased due to this non-Gaussian signal.

If one sorts the data sequence of interest (either in an ascending or descending order), the middle point, which is called the median is approximately the same as the mean value (theoretically they are the same, if it is a continuous infintely long Gaussian process) for signals from additive Gaussian noise channels. Furthermore, it gives a better estimate of the true mean even with interference involved.

Median filtering is, in a sense, a process that, if a finite number of data pass through the filter, the output will be the median value of the sequence. For an interference-contaminated Gaussian signal, the median filtering is superior to the direct averaging since the latter one no longer gives a good estimate, whereas the former one gives a more reasonable estimate of the true mean.

Interference can occur at any time, at any place, and in any form in the power spectrum data, a good process should be able to not only pin-point the interference but also correct it. An idea derived from the median filtering technique eventually leads to the solution of this problem. The process is described as follows.

Each spectral point in a power spectrum can be regarded as a random variable and the corresponding value is chosen from the parent population (this should refer to the infinite long sample space, but later on, we also use this terminology to designate the finite length sample space only for convenience). For a stationary process, and if no interference intervened, each parent population is Gaussian and has its own expectation value and variance. These expectation values then constitute the ideal or expected spectrum. Furthermore, if ergodicity applies here, each parent population can be generated by an infinite number of measurements in time. A three-dimensional probability density function of the D-region power spectrum from one height is shown in Figure 1 which explains this situation when the parent populations can be obtained and the process is stationary.

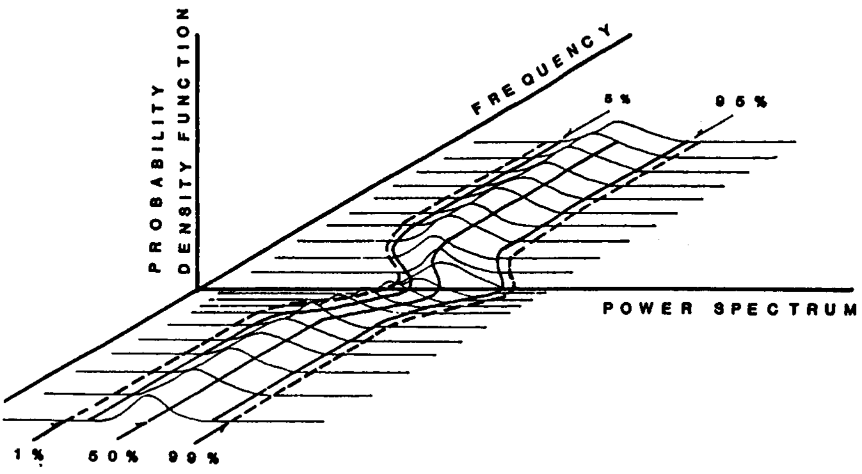


Figure 1. Three-dimensional diagram of the probability density function for each power spectrum point. Median value (50%), original template value (5%, 95%) and final template value (1%, 99%) are shown, respectively.

To form a statistically meaningful parent population in order to form a 'reasonable' median point, one has to include a large enough data set. In the case of D-region power spectrum measurement, the diurnal variation of the D region prohibits the time span to be too long, but one can incorporate the data from the neighboring heights within the same scale height, therefore, decreasing the number of the data in time to make the stationarity valid in the analysis.

Once a valid population for each spectral point of the power spectrum in question is formed, one can use the median filter technique and apply to each population created for each spectral point, then a series of median values, which is close to true mean, for every spectral point is obtained, this series of spectral points forms the median filtered power spectrum. But to carry out the template process, more information is needed in addition to the 'median' power spectrum during the median filtering process.

Template Formation. When the median filtering is carried out on a Gaussian-distributed data sequence, the middle point of this sequence is assigned as the median and that should be very close to the mean even if interference is present. In addition to the median, two extra outputs are obtained from the median filter. They are the values of the i^{th} point (i is less than N/2, the median point) and the (N-i^{th}) point and serves as auxiliary observation points of this data sequence for the template process. Here, N is the number of the sample points for each parent population and i is a number to be decided on next.

Obviously, the amplitude of interference should be always larger than the power spectrum itself. This is simply because the power spectrum data have no negative value and if the interference is smaller, then it will be hidden in the spectrum, so that no comparison can be made to this mixed signal whether it is interference or a real signal. If the median does point to the true mean of this data sequence and since the lower part of the power spectrum is almost not influenced by the interference, the lower margin value along with the median, bear very important information about the true statistics of this data sequence.

For example, if N is chosen as 100, then the values of the 51st point (or 50th point, since N is even in this case) is approximately the mean value of this Gaussian data sequence and if i is chosen as 5, then from the table (the table of the standard normal distribution function, LINDGREN, 1960), it is known that the 5th point stands for 1.645 standard deviation away from the mean. Since the Gaussian distribution is symmetric to the mean, one can use this information to set up a tolerance level. For instance, to cover 99% Gaussian data (2.33 times standard deviation), by taking the difference between the median value and lower 5% margin times 2.33/1.645 and adding the difference to the median value, the upper margin is formed. Consequently, all other spectral data outside these two margins can be treated as interference. Note that the interference may be included with greater chance if the 95th data point is used as the upper margin.

To be able to store and recreate these margins efficiently, a standard least-square Lorentzian fit was applied (the D-region power spectrum has the shape of Lorentzian distribution, see MATHEWS, 1984a,b; YING, 1985) to both the lower margin and the 'median' spectra and stores only four coefficients for each fit. Since almost no interference occurs at median and lower margin spectra (notice that this is the primary assumption we made to the template process), the Lorentzian fit gave out coefficients corresponding to the wanted signals, and when the templates are needed, these two 'fitted' power spectrum margins are used to decide the appropriate size of the template in order to detect and correct the interfered spectrum.

This is called template process only because this selecting procedure is just like putting a template onto the pdf-spectrum plot, as shown in Figure 1, and only those points within the template are selected. The template is shown as the region within two dashed lines in Figure 1. By using this template, the interferences not only can be detected but also can be modified in a sense that the good data and bad data are isolated in a single spectrum and then the good data can be processed instead of throwing out the whole spectrum.

Baseline Adjustment. One of the advantages of forming the templates is that the templates can be used to any spectrum within the 'height-time' windows that were used to generate the valid sample space for the median filtering process to remove the contaminated spectral points. But since the template formation process is done in a semi-stationary state because of the span in time and height baseline shift might exist throughout the experiment. Therefore, a proper baseline readjustment may be needed for certain spectra in order to have a proper 'cutoff' by the templates.

This is done by sorting the spectrum and assuming the median point is not interfered, then uses both the position and the value of the median point as references and aligns the center (average) of the two margin spectra at the corresponding position to this median point, consequently this spectrum should be positioned properly within the two margins of the template, and the proper 'cutoff' can be formed.

APPLICATION TO ARECIBO D-REGION POWER SPECTRUM

The power spectrum data that we are dealing with here were taken during a sequence of three and half days experiment from January 3 to 6 of 1981, at Arecibo, Puerto Rico. The basis of the experiment uses a 52 microsec 13-baud Barker coded pulse with 1 millisec interpulse period, which yields an effective height resolution of 0.6 km and 1 kHz bandwidth. For every 256 samples from the same height, a power spectrum was formed, results in a 3.9-Hz frequency resolution. A total of 63 heights spectra records were formed for each time designation.

The template formation process is carried out by using a 'window' of 20 records in time and 5 heights. This corresponds to a window that covers a time span of approximately one and a half hours, and 3 km height (note that 3 km height window is smaller than the nominal scale height of the D region which is around 5 km, therefore stationarity in height for the median filtering process can be secured). Since there are 63 heights, the highest process window includes height numbers from 59-63, overlaps with the next lower one, i.e., it shares the spectrum data of height numbers 59 and 60 with the second highest process window. The final template process on this experiment comprises 51 time spans, 13 different heights and thus 663 windows.

Figure 2 demonstrates a complete template formation process using the window from 11:46 to 13:13 of January 4, 1981, and height window 11 which comprises the actual height number from 51 to 55 and is about 89 km to 92 km in altitude. Figure 2a shows the actual spectra within the window with slight interference effect, each spectrum is populated into the plot and consequently there are 25600 spectral points in this figure. The interference can be seen from the upper part of the plot only; therefore, the lower part of the spectral points has reliable statistical significance. Figure 2b shows the median filtered result with the median spectrum and lower 5% spectrum superimposed onto the scattered spectral points. Apparently, the median spectrum is around the center as we expected. Figure 2c is based on Figure 2b with the fitted results displayed for the median and lower 5% spectrum. These two fitted spectra are used to form the two margins for the final template which is shown in Figure 2d as two dashed lines. 1% selection range was used in this case.

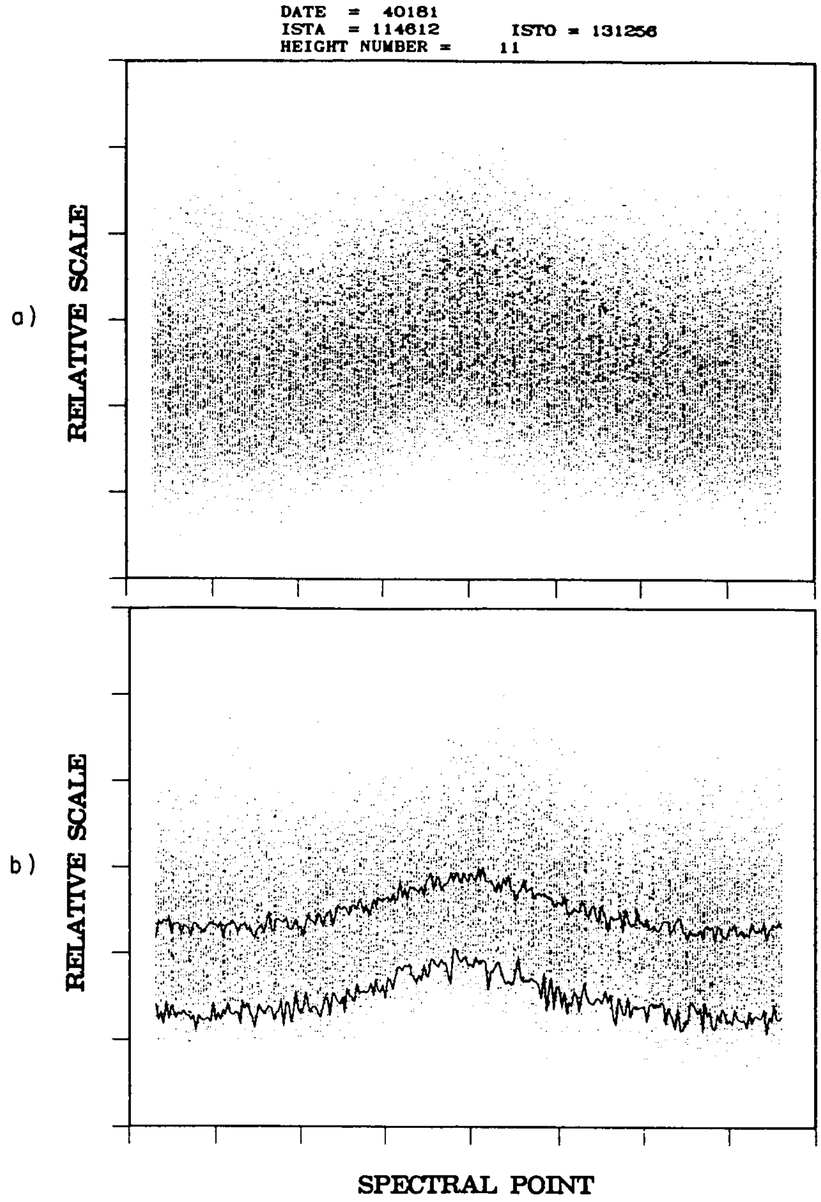
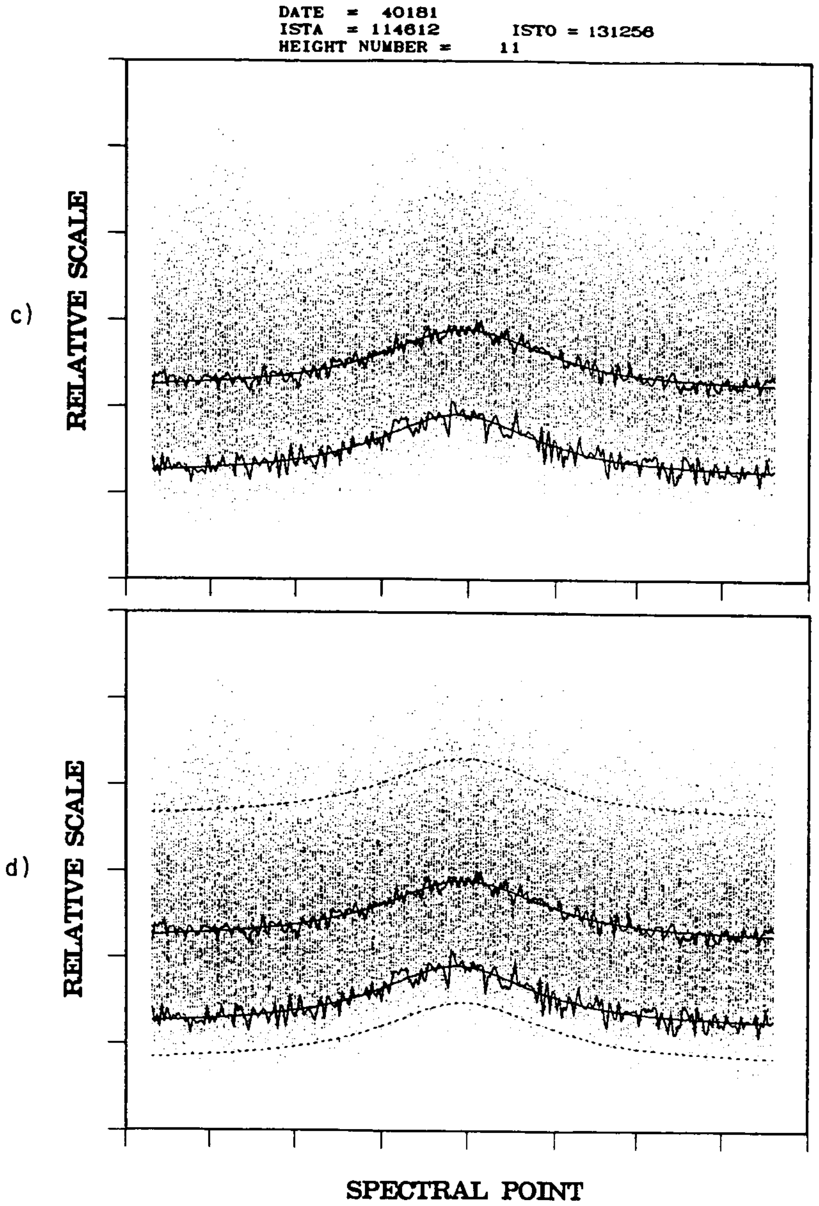


Figure 2. Demonstration of the template formation process for the time-height window from 11:46 to 13:13 January 4, 1981, in time and 89 km to 92 km in height. This is a slight interference contaminated case.
X axis represents spectral points from 1 to 256,
Y axis represents relative amplitude for all spectra.
(a). Spectral population of the specific window. There are 100 spectra within the window, altogether, 25600 data are present in this plot.
(b). Median filtering result of the specific window. The higher solid line represents the median spectrum, the lower solid line represents the 5% spectrum, the background is the spectral population.



(c). Fitted result of the two median filtered spectra, which are represented by two 'smoothed' solid curves. (d). The final templates formed by two margins shown in dashed lines. The template formed by two margins can be enlarged or reduced depending on how strictly the selection criterion is enforced. In this case, 1% selection is used.

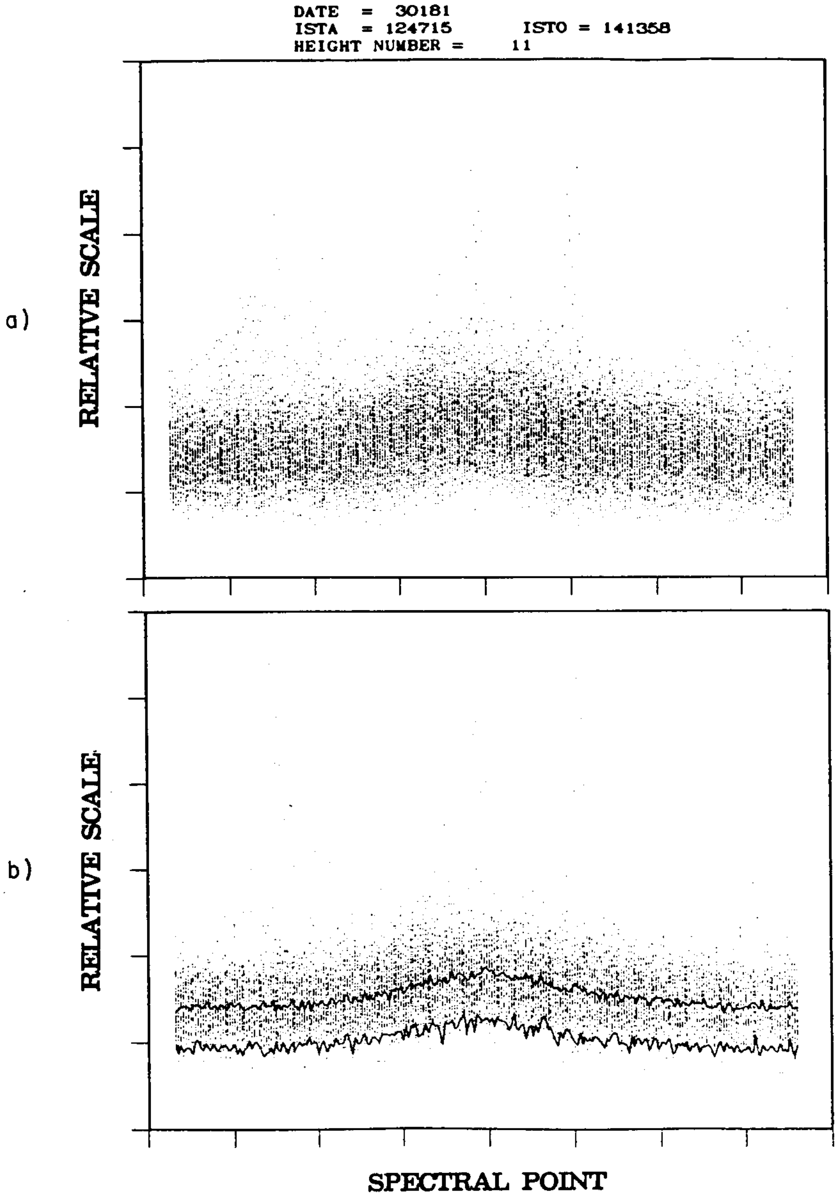
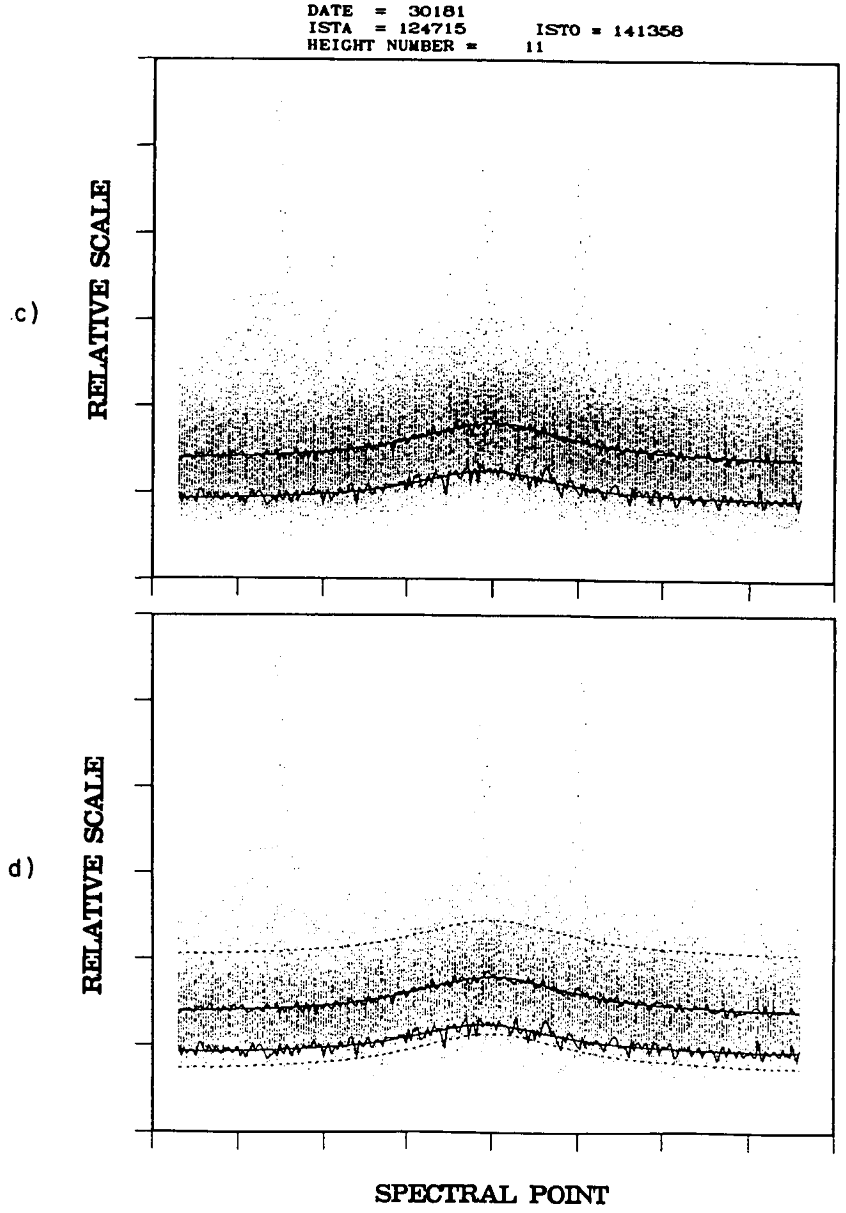


Figure 3. Demonstration of the template formation process for the time-height window from 12:47 to 14:14 January 3, 1981, in time and 89 km to 92 km in height. This is a severe interference contaminated case.
X axis represents spectral points from 1 to 256,
Y axis represents relative amplitude for all spectra.
(a). Spectral population of the specific window. There are 100 spectra within the window, altogether, 25600 data are present in this plot. Interferences can be easily seen from the upper part of this plot.
(b). Median filtering result of the specific window. The higher solid line represents the median spectrum, the lower solid line represents the 5% spectrum, the background is the spectral population.



(c). Fitted result of the two median filtered spectra, which are represented by two 'smoothed' solid curves. (d). The final templates formed by two margins shown in dashed lines. The template formed by two margins can be enlarged or reduced depending on how strict the selection criterion is enforced. For this 1% selection template, one can see how effectively the interferences can be detected and removed.

Figure 3 shows another template process for a window with severely interfered spectrum involved. The time is from 12:47 to 14:14 of January 3, 1981, the height window is the same as in Figure 2. Figure 3a shows such spectra with highly peaked interferences existing in the upper part of the window, but again the lower part of the window shows no evidence of 'outliers'. Figure 3b shows the median filtering process results and two more fitted spectra of both median spectrum and 5% spectrum are included in Figure 3c. Finally, Figure 3d reveals that the final template again successfully rejects the outliers and the remaining spectral points bear good statistics so that a further averaging and Lorentz fit can be applied.

SUMMARY

We have reported an almost universal interference detection and removal scheme which has been applied to the Arecibo D region power spectrum measurement to demonstrate the effectiveness of this method. The scheme comprises three major parts, each of them depends heavily on the stationarity of the process and the interference contaminating condition of the data.

The first part is the median filtering process which finds the median and 5% power spectrum from a collection of the spectra for which stationarity is assumed. Then the template is derived from the fitted median and lower margin spectrum. In this stage, and that not more than half of the spectra are contaminated is the major assumption here, so that both the median and lower margin spectrum can be assured to have the valid statistical meaning. The last process is to readjust the template to a proper position when performing the interference removal process, so that even though a slight nonstationarity exists during the template formation, a proper cutoff can be formed.

Another important idea has to be pointed out here is that, when further averaging is required of these interference-free spectra in order to gain the signal-to-noise ratio, one has to carry the template detection information along, so that a statistical weighting can be applied to each spectral point (YING, 1985) (for instance, the fewer points averaged together, the less statistical significance can be made to this averaged point). This is especially true when a further least-square fitting process is required, because the least-square fit depends on the statistics of each individual point that is fitted to.

ACKNOWLEDGEMENT

We wish to thank the staff of the Arecibo Observatory for their assistance, especially to Dr. M. P. Sulzer for sparing his time to evaluate this paper.

This work was supported by the National Science Foundation Atmospheric Sciences Division Grant ATM82-13619 to Case Western Reserve University. The Arecibo Observatory of the National Astronomy and Ionosphere Center is operated by Cornell University under contract with the National Science Foundation.

REFERENCES

Dougherty, J. P., and D. T. Farley (1963), A theory of incoherent scattering of radio waves by a plasma, 3, Scattering in a partly ionoized gas, J. Geophys. Res., 68, 5473.

Lindgren, B. W. (1960), Statistical Theory, The Macmillian Company.

Mathews, J. D. (1978), The effect of negative ions on collision--dominated Thomson scattering, J. Geophys. Res., 83, 505.

Mathews, J. D. (1984a), Incoherent scatter radar as a tool for studying the ionospheric D region, J. Atmos. Terr. Phys., 46, 975.

Mathews, J. D. (1984b), Incoherent scatter radar probing of the 60-100 km atm, Submitted to *IEEE Trans. on Geo-science & Remote Sensing*.

Rastogi, P. K. (1983), Data processing techniques used with MST radar: a review, unpublished paper, *Handbook for MAP*, *9*, 477.

Tanenbaum, B. S. (1968), Continuum theory of Thomson scattering, *Phys. Rev.*, *171*, 215.

Tepley, C. A., J. D. Mathews, and S. Ganguly (1981), Incoherent scatter radar studies of Mesospheric temperatures and collision frequencies at Arecibo, *J. Geophys. Res.*, *86*, 11330.

Ying, W. P. (1985), Analysis of the ionospheric ion line power spectra, M.S. thesis, Case Western Reserve University, Cleveland.