Radio Science, Volume ???, Number , Pages 1–36,

Techniques to Determine the Quiet Day Curve for a Long-Period of Subionospheric VLF Observations

Kathy Cresswell-Moorcock,[1] Craig J. Rodger,[1] Mark A. Clilverd,[2] David K. Milling,[3]

Corresponding author: K. Cresswell-Moorcock, Department of Physics, University of Otago, Dunedin 9016, New Zealand. (creal158@student.otago.ac.nz)

[1]Department of Physics, University of Otago, Dunedin, New Zealand.

[2]British Antarctic Survey (NERC), Cambridge, United Kingdom.

[3]Department of Physics, University of Alberta, Edmonton, Canada.

Very low frequency (VLF) transmissions propagating between the conducting Earth's surface and lower edge of the ionosphere have been used for decades to study the effect of space weather events on the upper atmosphere. The VLF response to these events can only be quantified by comparison of the observed signal to the estimated quiet-time or undisturbed signal levels, known as the quiet day curve (QDC). A common QDC calculation approach for periods of investigation of up to several weeks is to use observations made on quiet days close to the days of interest. This approach is invalid when conditions are not quiet around the days of interest. Longer term QDCs have also been created from specifically identified quiet days within the period and knowledge of propagation characteristics. This approach is time consuming, and can be subjective. We present three algorithmic techniques, which are based on either 1) a mean of previous days' observations, 2) Principal Component Analysis, or 3) the Fast Fourier Transform (FFT), to calculate the QDC for a long-period VLF dataset without identification of specific quiet days as a basis. We demonstrate the effectiveness of the techniques at identifying the true QDCs of synthetic datasets created to mimic patterns seen in actual VLF data including responses to space weather events. We find the most successful technique is to use a smoothing method, developed within the study, on the dataset and then use the developed FFT algorithm. This technique is then applied to multi-year datasets of actual VLF observations.

1. Introduction

Man-made VLF transmissions propagate for long distances with low attenuation in the Earth-Ionosphere waveguide between the conductive Earth surface and the lower edge of the ionosphere (D-region). This manner of propagation is termed subionospheric. The radiation propagates in a modal fashion within the waveguide, along the great circle path between transmitter and receiver, with the received amplitude and phase of the transmissions being a superposition of these modes [*Wait*, 1996a; *Lynne*, 2010]. Variations in the waveguide over time change the modal mix, causing variations in the observed signal amplitude at a receiver. The primary, temporally varying parameter of the waveguide is the D-region reflection height, which varies on a regular diurnal basis with the presence of solar radiation as well as irregularly in response to space weather events. For an overview of historical VLF science see *Barr et al.* [2000].

Diurnal variations in VLF observations have been used to determine the relationship of solar zenith angle to D-region parameters for the day-time ionosphere [*Thomson*, 1993], and determine similar parameters for the night-time ionosphere [*Thomson et al.*, 2007], when solar radiation has a less dominant influence. The most dramatic modal variations occur as the day-night terminator passes across the transmitter–receiver path, with the varying modal superposition causing very deep minima in signal amplitudes. *Clilverd et al.* [1999] related the timing of a set of twilight modal minima to the moving location of the terminator along a path at a height of 75 km. Lightning sferics within the VLF frequency band have also been used to determine day and night-time D-region parameters (e.g., [*Han and Cummer*, 2010a, b; *Shao et al.*, 2013]).

Space weather events, e.g., solar flares, energetic electron precipitation (EEP) from the radiation belts, and solar proton events (SPEs), can significantly increase the ionization rate in the upper atmosphere, which leads to a reduction in the D-region height and thus a perturbation in the received VLF amplitude [*Clilverd et al.*, 2009]. To study the effect of space weather events on the ionosphere using VLF propagation it is important to have a method of determining the undisturbed diurnal variation in the VLF observations, known as the Quiet Day Curve (QDC). Once the QDC is determined the perturbations caused to the VLF signals by these events can be quantified.

In this paper we report on the development of three algorithmic QDC finding techniques for long-period subionospheric VLF datasets. The first technique is based on the QDC technique from *Simon Wedlund et al.* [2014], who created their QDC from the combined curve of several quiet days prior to a period of geomagnetic disturbance that they were investigating. The second technique follows the QDC finding technique based on Principal Component Analysis reported by *Collier* [2009] and *Wautelet and Warnant* [2012]. These studies selected the principal components accounting for the most variance in their datasets for transforming back to data space. The third technique uses a 2-dimensional Fast Fourier Transform to identify the discrete spectrum of the dataset before applying restrictions developed within this study to the spectrum and inverting the transform to provide the QDC. These techniques all result in QDCs that reproduce both the diurnal and annual amplitude variations in the datasets, to some extent. A two-step pre-smoothing method, which was developed to improve the QDC technique results, is also described. Each technique is evaluated by its success at identifying the true QDC of synthetic datasets, which were created to mimic the behavior of real VLF datasets. The best technique is then applied to datasets of real amplitude observations of subionospherically propagated VLF and the results shown for example events. The development of a successful QDC finding algorithm for long-term datasets will allow for the detection of and statistical analysis of the ionospheric response to space weather events, observed through subionospheric VLF.

2. Datasets

2.1. AARDDVARK VLF Observations

The Antarctic-Arctic Radiation-belt (Dynamic) Deposition-VLF Atmospheric Research Konsortium (AARDDVARK) [*Clilverd et al.*, 2009] is a global network of VLF receivers located primarily in the polar regions [http://www.physics.otago.ac.nz/space/AARDDVARK_homepage.htm]. The AARDDVARK receivers providing data for this study are located near Scott Base (SB: 77°50′S, 16°39′E), Antarctica and Edmonton (EDM: 53°21′N, 112°58′W), Canada. The Scott Base receiver provides examples of long-distance transmitter-receiver great circle paths, while the Edmonton receiver provides examples of relatively short-distance paths. Observed signals from short and long paths are expected to behave differently due to the attenuation of higher level modes with distance [*Wait*, 1996b]. The datasets chosen from each receiver have the strongest signals of those monitored by those receivers and so provide relatively clear examples. Both receiver installations use UltraMSK software [*Clilverd et al.*, 2009] to process the VLF observations into amplitude and phase data. Figure 1a shows the locations of the two receivers, the monitored VLF communications transmitters that we utilize in our study, and the great circle paths between them.

In the current study we use only the VLF amplitude data, because there are extra difficulties in making long-term phase datasets consistent. An example is unpredictable phase jumps across periods when the transmitters are turned off, as discussed in *Rodger et al.* [2012]. We reduce the 0.2 s resolution raw data to 1 minute resolution by averaging, i.e., each data point in the reduced dataset is obtained by calculating the median of the 300 data points in that minute of raw data. This averaging removes much of the noisy variation inherent in the received signal from modulation of the transmission. By the law of large numbers, over the course of 1 minute this variation has a Gaussian distribution.

For the techniques described in this paper the data was arranged into a matrix such that each row contained one day's worth of data (1440 data points), with the rows ordered sequentially in time. Periods of abnormal transmission or interference from the receiver surroundings were removed from the dataset. Abnormal transmission periods were identified heuristically as when the signal dropped suddenly to the noise floor between 1 minute and the next and then later returned to normal signal levels. Periods of noise interference from the receiver surroundings were identified by comparison to observations on the 23.0 kHz frequency, which is rarely transmitted on.

Consistent temporal spacing of data points is essential for our QDC finding techniques, so periods when data was missing or removed were included as approximating values. The data approximations were done by combining the 'linear' method of the 'interp1q' MATLAB function for data gaps of >2 hr duration, the 'TriScatteredInterp' 2-dimensional surface interpolation function (using the dimensions of the data matrix) for data gaps 2 hr–2 days, and a median of the same numbered days of data from surrounding years for longer data gaps to maintain the overall coherence of the diurnal pattern within the approximated values.

Figure 1b shows 32 months of amplitude observations of the NDK (25.2 kHz) transmission received at the AARDDVARK antenna located near Edmonton, Canada. The great circle path between transmitter and receiver is completely dark in the primarily red region between 02 and 12 UT, and fully Sun-lit in the green to orange region between 14 and 24 UT. The border between the night and day regions is defined by the twilight modal minima, which vary their time of occurrence regularly through each year according to when the day-night terminator crosses the path.

2.2. Synthetic Dataset Creation

We created synthetic amplitude datasets for the purpose of evaluating the success of our QDC finding techniques at identifying the true underlying QDC of a dataset. These synthetic datasets were designed to be representative in their general response to light levels along a propagation path and to space weather events, rather than be a true model of the VLF dataset for the equivalent path. The synthetic dataset matrices contain four years of data at one minute resolution. Like the AARDDVARK data matrices, each row is one day of data arranged in UT time. Background patterns in the synthetic amplitude data simulate the general patterns seen in VLF amplitude data, with periods designated day-time (path fully Sun-lit), night-time (path fully dark) and twilight-time (day-night terminator along the path).

We present one of our synthetic datasets here, shown in Figure 1c, to illustrate the approach. Figure 1b shows AARDDVARK observations for the equivalent path: NDK-EDM (shown in Figure 1a). The day, night, and twilight-times of the synthetic dataset are defined by the solar zenith angles (SZA) at NDK and EDM. The diurnal variation in the synthetic dataset consists of four sections; a constant-valued section representing the VLF response to night-time conditions (approximately 02–12 UT), a curved section representing the VLF response to day-time conditions (approximately 14–24 UT), and two twilight sections separating the night and day-times, each with a single sinusoidal minima representing the twilight modal minima as seen in VLF data. The day-time curve ($Data_{day}$) is calculated as

$$Data_{day} = -(SZA_{NDK} + SZA_{EDM})/2 + 90$$

where SZA_{NDK} and SZA_{EDM} are the SZAs at NDK and EDM, respectively. A long-term trend of a single sinusoidal cycle is imposed on each column in the matrix. The diurnal variation is added to the long-term trend to form the background of the synthetic dataset, which is the true QDC that our techniques are aiming to identify. This background forms the dominant variation seen in Figure 1c. Perturbations are imposed, by addition, on the synthetic background to represent the VLF response to solar flares, EEP, and multi-day disturbances to the D-region. A fourth component imposed on the background represents the effect of random noise on the VLF signal. Figure 1d shows the background and combined data for a representative day from the synthetic dataset.

Across the 4 years of our synthetic dataset we impose 5000 “EEP events”, which we represent by downward pointing triangles, and 1000 “solar flare” events. The equation used to represent a solar flare event ($Flare$) is

$$Flare = 2x\, exp(-x/size)$$

where x is minutes from the start of the event, and $size$ is a random scale factor from 1 to 20, but biased to the lower end of the range. The imposed EEP events are placed only in the night-time region of the dataset, while the solar flare events are placed only in the day-time region of the dataset. The timing of both the EEP and the solar flare events is otherwise random, but biased towards periods of geomagnetic disturbance. While these events may not be strictly linked to geomagnetic disturbance, this bias gives a good representation of the clustering of space weather events which occurs in the “real world”. K_p index values for the four years spanning January 2009 to December 2012 (sourced from http://wdc.kugi.kyoto-u.ac.jp) provide a simple proxy for both solar and auroral activity and are used to supply the bias, where a higher K_p will lead to more imposed synthetic EEP and solar flare perturbations. The magnitude of the imposed EEP and solar flare events is randomly generated within the range 0.6–15 dB, which is representative of the range of responses caused by solar flares and EEP seen in real VLF datasets.

Multi-day perturbations are included to simulate the effect of longer space weather events, such as SPEs, or longer-term geomagnetic disturbances. In our synthetic datasets the timing and strengths of these perturbations in the dataset are determined by the D_{st} index values (sourced from http://wdc.kugi.kyoto-u.ac.jp) for the same 4 year period as used for the K_p-based perturbations. The range of D_{st} values in the period was divided into disturbance levels, which were used to assign perturbation values to each entry in the synthetic data matrix. These added values were smoothed to remove sharp steps from the perturbations. The magnitude range of the added values is 0–5 dB, negative during night-time and positive during day-time. We placed no restrictions on the length of the multi-day perturbations, beyond those inherent in the D_{st} dataset disturbance levels.

The added noise component consists of random values selected from a zero-centered Gaussian distribution in the range $\pm x$ that are added to each data point in the day-time and night-time sections of the dataset. We define x from the uncertainties reported by *Rodger et al.* [2007]. The distribution standard deviation during day-time is 0.02 to give an x of 0.1 dB and during night-time is 0.1 to give an x of 0.5 dB.

3. Technique Descriptions

Below we give descriptions of the QDC finding techniques developed in this study. We also describe the pre-processing addition that we developed to improve the results of the techniques. In the development of these algorithms, we have aimed to make them generic and not specific to one known dataset. As such, these approaches should be valid for any subionospheric VLF amplitude dataset of sufficient duration.

3.1. Combined Daily Curve

This technique generalizes the method used by *Simon Wedlund et al.* [2014]. They calculated their QDC from the combined curve of several identified quiet days of VLF amplitude observations that occurred shortly before a period of geomagnetic disturbance. In the current study this method is generalized by applying the technique with no regard for the level of disturbance in the previous days' data, i.e., there is no attempt to determine if the previous days are indeed quiet. This is done so that our technique does not rely on the time-consuming manual identification of quiet days within a dataset. We therefore note that the calculated QDC will be of lower quality than if we knew the utilized observations came from a truly quiet period. Thus, this technique may best suit periods of lower solar activity. We refer to this method as a Combined Daily Curve (CDC).

The CDC is created by averaging data from the 3 days prior to the day of interest. The CDC technique assumes that the diurnal pattern in VLF data changes very little from day to day, except in response to ionospheric perturbations, which the averaging is expected to remove. This assumption is based on examination of diurnal patterns in VLF datasets (e.g., the relatively regular variations seen in Figure 1b).

The CDC is calculated at 10 minute resolution, with each value in the CDC being averaged from the same respective 10 data points in each of the previous three days. Thus, each average value is calculated from thirty 1 minute data points to match the thirty data points, from a single day, that were used by *Simon Wedlund et al.* [2014] for their QDC value calculation. The CDC is then interpolated back to 1 minute resolution, using the MATLAB function 'interp1' with the 'linear' method, for direct comparison with the data. We change the resolution in this manner to reduce the influence of any one data point on the result. The first 3 days of data in the matrix do not have corresponding CDCs as they do not have at least 3 days prior to them.

3.2. Principal Component Analysis

Principal Component Analysis (PCA) is a tool used in multivariate analysis to expand a dataset along its directions of maximal variance. For analysis of data variation, this expansion is sufficient. However, it is possible to summarize the patterns in a dataset by selecting expansions along a limited number of directions of highest variance and recombine them [*Collier*, 2009]. For the purpose of this QDC finding technique, we assume that the majority of the variance in the dataset comes from the regular diurnal patterns of the data and is thus concentrated in the lower ordered PCA directions.

The steps of the PCA QDC finding technique for an $m \times n$ data matrix $\mathbf{X}$ are as follows.

1. Create the re-centering matrix $\overline{\mathbf{x}}$, which has the entries of each column as the mean of the corresponding column of $\mathbf{X}$.

2. Calculate the covariance matrix $\mathbf{S}$, of the recentered data matrix.

$$\mathbf{S} = \frac{1}{m-1}(\mathbf{X} - \overline{\mathbf{x}})'(\mathbf{X} - \overline{\mathbf{x}}).$$

3. Find the eigenvectors and eigenvalues of $\mathbf{S}$. These should be sorted in decreasing order by the eigenvalues. The eigenvectors are the directions of maximal variance for the PCA process and the corresponding eigenvalues give the variance accounted for by each direction.

4. Project the recentered data matrix onto the eigenvectors of $\mathbf{S}$ to find the principal components (**PC**s). Defining $\mathbf{G}$ as the matrix of eigenvectors, arranged column-wise, the matrix of principal components $\mathbf{Y}$, is

$$\mathbf{Y} = (\mathbf{X} - \overline{\mathbf{x}})\mathbf{G}.$$

Each column of $\mathbf{Y}$ is a single PC. The PCs are ordered according to the variance accounted for by their corresponding directions, with the first being the projection of the recentered data matrix onto the direction of highest variance.

5. Choose and apply the criteria to be used for limiting the number of PCs. We use the Kaiser criterion [*Kaiser*, 1960], which retains only those PCs that individually account for more than the mean variance over all the PCs.

6. Invert the projection for all retained PCs, sum them together and add the re-centering matrix. With $\mathbf{y}_{(1,2,\ldots i)}$ and $\mathbf{g}_{(1,2,\ldots i)}$ defined as containing the retained PCs

and corresponding eigenvectors respectively, the resulting QDC matrix $\mathbf{Q}_{PCA}$, is

$$\mathbf{Q}_{PCA} = \mathbf{y}_{(1,2,\ldots i)}\mathbf{g'}_{(1,2,\ldots i)} + \overline{\mathbf{x}},$$

3.3. Fast Fourier Transform

The Fast Fourier Transform (FFT) is used to identify the discrete frequency spectrum of a digital dataset. In this study the two dimensions of the FFT are the diurnal variation in the rows of the data matrix and the day-to-day variation, which includes the yearly variation, in the columns of the data matrix. Our FFT QDC finding technique uses the 2-dimensional transform to calculate the spectrum of a dataset, which is then restricted as described below. We calculate the inverse transform of the restricted spectrum to provide our QDC. *Amidror* [2013] gives an overview of the transform in multiple dimensions including details of various issues to be aware of when using the transform.

In this technique we want to remove as much of the perturbation contribution from the spectrum as possible while retaining as much of the background contribution as possible, as this represents the true QDC we are trying to find. The central aspect of this technique is the identification of the spectral components that are dominated by the perturbation spectrum. Once these unwanted components are identified, we remove their contribution to the spectrum by setting them to zero. The QDC is taken as the real component of the resulting matrix from the inverse FFT. Note that providing the spectrum restrictions maintain the symmetry properties of the original spectrum, the result of the inverse FFT will have no imaginary component.

The linear property of the FFT allows for the examination of the features of the synthetic background spectrum independently from the perturbation spectrum. From this examination we are able to identify consistent features of these spectra across multiple synthetic datasets with different backgrounds and thus develop methods to identify perturbation-dominated spectral components for removal from the spectra.

The first spectral restriction is the removal of certain rows of the FFT spectrum to clarify the yearly, including seasonal, variation of the dataset. For this clarification to be most effective, the dataset is required to be a whole number of years, say p, in length. Cutting the dataset prior to application of the FFT may be required to achieve this. The yearly background pattern of a p-years length dataset repeats p times in the vertical direction of the data matrix. This regular repetition places the background-related spectral components on the p^{th}-multiple vertical frequencies, or rows from the center, of the spectrum. Spectral leakage is a frequency smearing artifact in the FFT that results from the effective discrete truncation of a continuous function [*Amidror*, 2013]. It causes all spectral components in the spectrum to contribute to those surrounding them, in this case the result is that the non-p-multiple rows of the spectrum have some contribution from the background patterns. By limiting the dataset to whole numbers of years we minimize that contribution, allowing us to assume that the non-p-multiple rows are perturbation-dominated. Thus, by keeping the dataset to p years, we can immediately identify the non-p-multiple rows of the spectrum as being perturbation-dominated and set their components to zero for QDC generation.

This first spectral restriction essentially requires datasets to be of longer duration than two years to allow for row removal in the spectrum. Due to this requirement, our FFT QDC finding technique is not valid for VLF datasets shorter than 2 years.

The second spectral restriction is the removal of two regions of the spectrum matrix that are consistently perturbation-dominated and are located vertically up and down from the center of the matrix, and the retention of background-dominated regions. Separate examination of background and perturbation spectra from our synthetic datasets showed us the regions in the combined spectra where each would be expected to be dominant. The strong spectral components of the background layers are located in the center of their spectra, fanning outwards horizontally and diagonally with decreasing magnitudes in patterns specific to each background. Figure 2a shows the spectral magnitudes of the central section of the synthetic spectrum. Here the background-related pattern is seen on every 4^{th} row as a higher magnitude than surrounding values. None of the background spectra fan out in the vertical directions. The strongest spectral components of the perturbation layers are located in the central column of their spectra (the vertical green columnar region in Figure 2a), symmetrically reducing in magnitude with horizontal distance. From these observations we find that the two triangular regions located in the vertical directions from the center of the spectrum have little contribution from the background spectra and are thus perturbation-dominated. The boundaries of the region of strong background-related spectral components are different for each background and must be identified separately for each dataset. Once the boundaries of the region of significant background contribution are identified, the spectral components in the triangular regions outside of the boundaries are easily set to zero using a stencil.

The third spectral restriction is the removal of low-energy spectral components in the high frequency regions of the spectrum matrix. At the edge of the spectrum matrix, where the frequencies are highest, the spectral components are perturbation-dominated and the spectral magnitudes are relatively low. It is necessary to identify the border of the matrix region within which the background-related spectral components are dominant. This is the point where the distinct pattern of the background-dominated spectral components is subsumed into the general spectrum. A spectral energy limit is employed, with the limit chosen as the lowest energy at which the background pattern is retained and a minimum of spectral components from outside of the pattern are included. This method is less subjective than a determination through visual inspection to find the border of the background-dominated region of the spectrum. The spectral energy limit is different for each spectrum due to the differing background patterns in each corresponding dataset. For the chosen energy limit, a plot of the inverse FFT of the discarded spectral components should not include background patterns from the dataset or periodic variations of greater than 0.1 dB magnitude.

The first spectral restriction tends to remove contributions from long-term trends to the spectrum of the dataset, due to VLF dataset long-term trends likely being a response to the solar activity cycle of 11 and 22 years. Unless the dataset is itself a multiple of 11 years in length, the main trend-related components are lost at the row removal stage. Thus for this QDC finding technique to take into account any long-term trends, an extra step is needed to re-include the strongest of the removed spectral components in the low frequency region of the matrix to the spectrum prior to the inverse transform.

The final synthetic spectrum, after all the restrictions have been applied, is shown in Figure 2b. As with Figure 2a we show only the spectral magnitudes from the central section of the spectrum. The combination of the row removal and stencil restrictions has removed the visible contribution of the perturbation-dominated components in the central region of the spectrum, while the removal of lower energy components shows the border of the background-dominated matrix region.

3.4. Additional Smoothing

As will be reported in Section 4.2, the three basic QDC finding techniques, described above, produce promising results when applied to our synthetic datasets. We also investigated methods to pre-smooth the datasets with the aim of improving the results from the basic techniques. We found that a two-step pre-processing approach, which involves the removal of the most disturbed days of data and then a smoothing of the resulting matrix, applied to the dataset prior to application of the QDC finding technique provided an improvement in the results for the day-time and night-time regions of the matrix. These pre-processing methods are described below.

The results of all three QDC finding techniques are negatively influenced by periods of significant disturbance in the datasets to some degree. We investigated nearest neighbor distances [*Cover and Hart*, 1967] as a method of defining the disturbance level of a row of data. Figure 3a shows the nearest neighbor distance for each row of the synthetic perturbation matrix plotted against the nearest neighbor distance for the corresponding rows of the full synthetic dataset. Here we see that rows with higher dataset distances also have higher perturbation distances. From this relationship, we determine that the dataset nearest neighbor distance of a row is a good indicator for the actual disturbance level of a row. We therefore remove from the data matrix those rows with the highest 10 % nearest neighbor distances, as the most disturbed. In Figure 3a this limit is marked by a dashed vertical line.

We then smooth the data, which serves two purposes: to replace the removed data from disturbed days and reduce the influence of short term perturbations, i.e., solar flares, on the QDC. We use the 'rloess' method of the 'smooth' function from the MATLAB® software package's Curve Fitting Toolbox. This method is a "local regression using weighted linear least squares and a 2nd degree polynomial model" that "assigns lower weight to outliers in the regression" [www.mathworks.com/help/curvefit/smooth.html]. The 'rloess' method was preferred for the smoothing over a moving average, because of the lower influence of outlying values on the result under this method. This smoothing method fills gaps in the input data as part of the algorithm. We found that smoothing over the gaps from the removed disturbed days in the data matrix improves our results even more than filling them with representative values. The long-term trend in the data is not significantly affected by this method of smoothing as shown by the daily data means presented in Figure 3b. The smoothed dataset daily means show significantly less variation than those of the unsmoothed, full, dataset while also remaining close to the background daily means.

The smoothing is done both column-wise and row-wise in the data matrix. The column-wise smoothing is intended to remove single day perturbations, which can be considered outliers within the general shape of the data from day-to-day, and mitigate the effect of multi-day perturbations, such as SPEs. The row-wise smoothing is intended to further reduce the effect of noise around the signal.

Care must be taken in choosing the span for the smoothing. Too high a span and the desired background patterns in the data are lost, too low and the smoothing is practically pointless. We tested a range of spans on various of our synthetic datasets to determine the level required under these constraints. For the twilight-times, we found that a span of 7 data points provides adequate smoothing of perturbations without significantly altering the shape of the minima. A higher span is possible for the day-time and night-time regions of the data matrix. We found that a span of 13 data points provided very good smoothing while limiting the addition of negative artifacts to the smoothed data matrix in these regions. We therefore smooth the data matrix twice, once at a span of 7 and once at 13, and combine the twilight-time region of the 7-span result with the day and night-time regions of the 13-span result to give our final smoothed dataset for application of a QDC finding technique.

A low-pass filter might be used here as an alternative to the smoothing. However, it is not clear whether this style of filter would provide a significant enough improvement to the results of the method described above to justify the added subjectivity of determining the cut-off frequency for each dataset. Our smoothing method is convenient to the MATLAB® user and requires little subjectivity in the identification of the required span, which can then be easily translated across different datasets.

4. Testing Techniques on Synthetic Data

4.1. Method to Quantify Technique Success

We evaluate the success of our QDC techniques by calculating a parameter to indicate how close our QDC matrices are to the synthetic background, which is the true QDC of the synthetic dataset. This parameter allows us to directly compare the success of our techniques. We calculate this parameter from the difference between the QDC and the background, which we refer to as the Comparison. Clearly, it is only possible to determine this parameter for synthetic datasets due to the true background being unknown for real VLF observations.

Our indicative parameter is based on the L^2 vector norm and so we will refer to it as the norm for the remainder of this study. The equation used to define the norm is

$$||\mathbf{v}|| = \sqrt{\sum_i v_i^2/n}$$

where $||\mathbf{v}||$ is the norm, v_i are the entries in the relevant section of the Comparison matrix and n is the number of entries in the section. The norm parameter is higher than a simple average of absolute values due to the squaring of the entries. It has no direct physical meaning, being used here as an estimation of the outer variability of the Comparison matrix. The norm can be calculated for each section of the Comparison matrix, night-time, day-time, and twilight-time, as well as for the complete matrix. This allows us to compare technique success between Comparison sections.

For our technique evaluation we use ten different synthetic datasets, with identical backgrounds, that differ only in the random timing and magnitude of the imposed perturbations. The final reported norms, in Table 1, for each technique are the mean of the ten norms found for the application of the specific technique to each of the ten datasets. The uncertainty is taken as the range of the norms over the ten datasets and is also reported in Table 1.

Table 1 has two sections, with the norms of the upper section for application of the QDC finding techniques (outlined in Sections 3.1–3.3) to the synthetic data, and the norms of the lower section for the inclusion of the two step pre-smoothing method (outlined in Section 3.4) prior to application of the techniques. The norms in each section of the Table are arranged by technique and region of the Comparison matrix: 'All' for the entire synthetic dataset, 'Day' for periods when the path is fully Sun-lit, 'Night' for the periods when the path is fully dark, and 'Twilight' for the periods when the day-night terminator intersects the path.

Lower norms result from technique calculated QDCs that are closer to the synthetic background, on average. Thus the best technique is the one resulting in the lowest norms. The norms in the top row of Table 1 compare the complete synthetic dataset, including all the imposed perturbations, to it's background. These norms are the absolute upper boundary of what we would accept for the results from a QDC technique as a higher norm would imply extra perturbations have been added by a technique.

The following subsections give the quantitative evaluation of the 'Basic' and 'Pre-Smoothing' techniques by their norms, as displayed in Table 1. Qualitative evaluation is provided for each technique by Comparison plots, i.e., the difference between the calculated and true QDCs. These plots are given in Figures 4 and 5 and are each processed from the same representative dataset of the ten used in the testing.

4.2. Evaluation of Basic Techniques

Comparing the norms within the upper section of the Table, we see that in the Twilight sectors the CDC and PCA QDCs result in higher norms than those for the synthetic dataset itself. In the Day and Night sectors, all three QDC techniques result in lower norms than those of the dataset. Across all sectors the FFT QDC finding technique shows the best results, with norms of less than 1 dB, whereas the CDC and PCA techniques both result in norms greater than 1 dB.

Figure 4 shows the Comparison plots for the synthetic data and all three basic QDC techniques. Plot (a), Data, is effectively just showing the synthetic perturbations, as expected. Plot (b) and plot (c), for the Basic CDC and Basic PCA techniques, respectively, show significant remaining influence of the imposed multi-day perturbations. Plot (d), Basic FFT, shows less localized influence of the perturbations than is seen in the plots for the other techniques. However, the overall effect of the imposed perturbations for this technique is to bias the calculated QDC in the dominant direction of the data disturbance, either positive (day-time for the synthetic datasets) or negative (night-time for the synthetic datasets).

The norms and Comparison plots for the three basic QDC finding techniques indicate that the basic FFT technique is promising, but has the significant issue of bias, which will be important in practical application. However, further investigations found that these results can be significantly improved upon and the next section gives the analysis for the addition of the developed pre-smoothing method to the techniques.

4.3. Evaluation of Pre-Smoothing Techniques

The complete algorithm for each technique evaluated in this subsection involves applying the two step pre-smoothing method, described in Section 3.4, to the full synthetic dataset and then applying the chosen QDC finding technique to the resulting data matrix.

The top row of the lower section of Table 1 gives the norms for the comparison of the smoothed synthetic dataset to the background. Here we see an immediate improvement over all of the Basic norms in the upper section of the Table, excepting only the Twilight norm for the FFT technique.

Applying either of the CDC and PCA techniques to the smoothed synthetic data gives no improvement to the norms over the smoothing alone. Applying the FFT technique to the smoothed data improves the results in all sectors, almost halving the norms from the smoothing alone. The day-time norm for the pre-smoothed FFT technique is 0.23 dB, which is around twice the maximum level of the day-time imposed noise (0.1 dB). The night-time norm is 0.14 dB, which is less than half of the maximum level of the imposed night-time noise (0.5 dB). In contrast, the norm for the twilight-time section has increased compared with that of the basic FFT technique.

Figure 5 shows the Comparison plots for the smoothing method and pre-smoothed QDC finding techniques. Note the color scale range of this figure has been decreased from Figure 4. Plot (a), Smoothed Data, shows significant removal of perturbations from the calculated QDC, with only localized influence of highly perturbed periods in the synthetic dataset. Plot (b), Pre-smoothed CDC, shows no improvement over the Smoothed plot during the times of highly perturbed periods. The yellow and blue regions between 12 and 24 UT in the CDC plot show that a simple average of previous days as a QDC is prone to influence from any day-to-day slope present in the data, i.e., during Sun-lit periods in the synthetic dataset (14–24 UT, Figure 1c). Plot (c), Pre-smoothed PCA, shows the difficulty of separating background-related variance from perturbation-related variance in the PCA process. In this plot, the vertical sections encompassing the periods of twilight modal minima (23–04 UT and 11–16 UT) show a distinct lack of definition for the minima while other sections are clearly influenced by the perturbations remaining in the smoothed synthetic data, such that they appear in our calculated QDC. The PCA QDC finding technique may have more success at identifying the true QDC for a shorter period dataset, of maybe month duration, however, investigation of this possibility is beyond the scope of this study. Plot (d), Presmoothed FFT, still shows some bias in the calculated QDC to the dominant direction of the data disturbance, however, this bias has been significantly reduced from that seen in the Basic FFT Comparison plot of Figure 4. While the pre-smoothed FFT technique does not represent the modal minimum periods well, in general this technique provides the best calculated QDCs.

We conclude from the norms presented in Table 1, and examination of the plots in Figures 4 and 5, that the best of the methods considered in this study for identifying a QDC of a long-lasting VLF dataset, is to smooth the dataset as described in Section 3.4 then apply the FFT technique as described in Section 3.3. Unfortunately the restriction of the FFT technique to datasets of at least two years duration, to allow the row removal step to be applied, means that this technique is not appropriate for shorter datasets. Thus, for datasets of less than two years duration we recommend the pre-smoothing process alone as the best method for identifying a QDC.

Figure 6 shows a single representative day of synthetic data and the results for the pre-smoothing process and the FFT QDC finding techniques. 6a is the synthetic data and the two QDC results, which follow the diurnal pattern in the data visually successfully. 6b shows the imposed perturbations for the day and the difference between the data and each QDC, which we call the Remainder. At the visual level, the Remainders contain the imposed perturbations. 6c shows the Comparison, which is the difference between the true and calculated QDCs or equivalently between the perturbations and the Remainder, for the two QDC results. For this day, the FFT has larger magnitude Comparison values than the pre-smoothing process does in general, however both lines on the plot remain within 0.25 dB of zero for most of the day.

We note that while the described methods give good results for identifying the QDC from perturbations occurring during relatively slowly changing sections of data, such as is usually seen when the VLF path is either fully Sun-lit or fully dark, the sharp amplitude changes seen around the twilight modal minima times are not so well dealt with. At this point we struggle to produce an accurate QDC representing the intensity of twilight-time amplitude variations. Therefore caution is advised in the interpretation of QDC finding technique results around the times of twilight modal minima.

5. Application to actual AARDDVARK Datasets

We now provide example results of the application of this overall technique to our AARDDVARK VLF datasets. We take the smoothing spans that were used for the synthetic datasets and use these spans for the smoothing of the AARDDVARK VLF datasets.

5.1. Clarifying the FFT Spectrum

When we began applying our FFT QDC finding technique to real VLF observations, we found that the background-dominated central pattern of the FFT spectrum was less distinct for some datasets than for the synthetic dataset. This lack of clarity of the central pattern was identified as being caused by two sources. Firstly, the dynamic range of amplitudes for a VLF dataset is usually much less, varying from 42 to 55 dB for the datasets used in this study, than the approximately 100 dB used for the synthetic dataset. That value was set to ensure clear diurnal variations rather than as an actual model of real VLF data. Secondly, the twilight-time modal minima patterns in the synthetic dataset were based on a relatively short transmitter–receiver path (NDK to EDM in Figure 1a at 1.304 Mm) and so had a very simple structure, which made the background-related spectral patterns clear in the overall spectrum. Longer paths demonstrate more complex twilight modal interference patterns due to there being more distance along the path for interference fringes to occur [*Clilverd et al.*, 1999]. The background-related spectral patterns in the spectrum are less clear as the path lengthens, such as for the three Scott Base recorded transmitters in this study.

In order for the parameters of the restriction stencil to be correctly identified when the amplitude dynamic range is small and the modal interference patterns in the dataset complex, the central pattern of the real VLF spectrum needs to be clarified. We do this by subtracting an average magnitude row (found from the perturbation-dominated higher frequency region of the spectrum) from the magnitudes of each row of the overall spectrum, which leaves an approximate indication of the background-related pattern in the spectrum for identification of the stencil boundaries. The stencil is then applied to the "unclarified" spectrum as normal. With this addition to the FFT QDC finding technique, the response of the real VLF datasets to the technique improves.

5.2. Application Results

Figure 7 shows the dataset, calculated QDC, and the difference between the two (Remainder) for 5 years of amplitude observations for the NWC (19.8 kHz) transmission received by the AARDDVARK antenna near Scott Base, Antarctica. The overall background patterns of the dataset appear well reproduced in the QDC. However, as the true QDC for real VLF amplitude observations is unknown, this is impossible to quantify. Some of the modal minima regions of the Remainder plot still show consistent amplitude differences, in contrast to the day-time and night-time regions, where the differences appear dominated by true perturbations.

Details from the Remainders in Figures 7c and 6b suggest that our FFT QDC finding technique is successful at identifying VLF responses to solar flares. This is confirmed for real VLF observations by examples of the VLF Remainder response to solar flares shown in Figure 8. These plots also show the, flare-defining, GOES satellite observed solar X-ray (0.1–0.8 nm) flux for the same period. The NWC-SB path was partly-lit until approximately 21:30 UT when it became fully Sun-lit, but still shows a visible response to the M1.7 flare, which occurs during the period of partial illumination. The other four paths were fully Sun-lit during the times of the shown solar flares. Variations in the solar X-ray observations outside of the flares are also seen as variations in the NLK–SB and NPM–SB observations. These examples demonstrate our QDC-finding technique's success at identifying the underlying variation for relatively short-duration space weather events.

Figure 9 shows an example of a VLF response to a SPE for the NLK transmission observed by the Scott Base receiver. An SPE is defined for space weather purposes by the proton flux at energies >10 MeV exceeding a threshold of 10 (cm^2 s sr)$^{-1}$ at geosynchronous orbit. The QDC in 9a shows a consistent diurnal variation, which the amplitude data largely follows before the SPE begins and after the SPE flux has returned to relatively quiet levels, i.e., approximately 80–96 hours in the plot. 9b shows only the Day-time and Night-time Remainder. We do not show the Remainder for the twilight modal minima periods in accordance with the caution advised for the interpretation of the QDC during these periods. The Day-time Remainder shows a clear offset from zero for the first two periods when the VLF path is Sun-lit after the SPE begins. The Night-time Remainder shows a general offset from zero for the first three periods after the SPE begins, although with more variability than the Day-time periods show. Note that the SPE is clearly still affecting the data in the third Night-time period even though the SPE flux is below the SPE threshold for this period. 9c shows the corrected >10 MeV Proton flux observations from GOES-13 for context. The VLF amplitude response to changes in waveguide parameters varies depending on the result of the superposition of multiple propagating modes. This will not generally lead to a linear relationship between the perturbing SPE flux and the observed remainder, as this figure shows. The remainder here demonstrates our QDC-finding technique's success at identifying the underlying variation even during space weather events lasting multiple days. This figure also shows that the D-region exhibits sensitivity to solar protons for fluxes below the SPE threshold.

6. Summary and Conclusions

In this paper we described three algorithmic techniques for the calculation of Quiet Day Curves for observations of VLF transmissions propagated subionospherically.

1. The Combined Daily Curve technique calculated an average of the previous three days' data for its QDC.

2. The Principal Component Analysis technique transformed the data matrix to the directions of maximal variance, selected those directions accounting for more than the mean variance and transformed them back to data-space for its QDC.

3. The Fast Fourier Transform technique transformed the data matrix to its discrete spectrum, restricted those spectral components likely to be perturbation-dominated, and transformed the restricted spectrum back to data-space for its QDC.

In addition, a smoothing process was described for application to the data prior to a QDC finding technique.

We evaluated the success of these techniques at identifying the true QDCs of perturbed synthetic datasets and identified the algorithm combining the pre-smoothing process (described in Section 3.4) and the Fast Fourier Transform based QDC finding technique (Section 3.3) as the most successful technique on average over an entire dataset. This combined technique was found to identify the true QDC of our synthetic datasets to within 0.23±.02 dB during Day-defined periods and within 0.14±.01 dB during Night-defined periods of the datasets. The fast modal variations during the Twilight-defined periods were identified to within 0.77±.05 dB. The FFT based technique is only valid for datasets of at least two years, for shorter datasets the pre-smoothing process alone, which was found to give the second best results in the evaluation, is recommended as a QDC finding technique.

The combined pre-smoothing and FFT based QDC finding technique was then applied to real datasets of observed VLF transmissions, from the AARDDVARK receivers located near Scott Base and Edmonton. Example results for five transmitter-receiver paths were provided to demonstrate the technique's ability to identify responses to perturbations across the entire dataset (Figure 7), to solar flares (Figure 8), and to a multiple day SPE in real-world VLF data (Figure 9). From these examples we deduce that this FFT based QDC finding technique will allow for statistical analysis of VLF responses to space weather events occurring in datasets of longer duration than 2 years.

Acknowledgments. The synthetic dataset used throughout this paper and the file used to create it are available from the corresponding author in MATLAB .mat and .m formats respectively.

AARRDVARK VLF data availability is described at its website:
http://www.physics.otago.ac.nz/space/AARDDVARK_homepage.htm

The GOES-13 proton (>10 MeV) corrected flux data used in Figure 9c was downloaded (02/12/2014) from online file: http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/2011/11/goes13/csv/g13_epead_cpflux_5m_20111101_20111130.csv

The GOES-14 X-ray data used in Figure 8a was downloaded (28/06/2014) from online file: http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/2010/01/goes14/csv/g14_xrs_1m_20100101_20100131.csv

The GOES-15 X-ray data used in Figure 8b was downloaded (12/11/2014) from online file: http://satdat.ngdc.noaa.gov/sem/goes/data/new_avg/2010/01/goes14/csv/

g15_xrs_1m_20111101_20111130.csv

KCM and CJR would like to acknowledge support from the New Zealand Marsden Fund.

KCM would like to acknowledge support from a University of Otago Publishing Bursary.

MAC would like to acknowledge support from the Natural Environmental Research Council grant NE/J008125/1.

References

Amidror, I. (2013), *Mastering the Discrete Fourier Transform in One Two or Several Dimensions - Pitfalls and Artifacts*, Springer-Verlag.

Barr, R., D. L. Jones, and C. J. Rodger (2000), ELF and VLF radio waves, *Journal of Atmospheric and Solar-Terrestrial Physics*, *62*(17-18), 1689–1718, doi: 10.1016/S1364-6826(00)00121-8.

Clilverd, M. A., N. R. Thomson, and C. J. Rodger (1999), Sunrise effects on VLF signals propagating over a long north-south path, *Radio Science*, *34*(4), 939–948, doi:10.1029/1999RS900052.

Clilverd, M. A., C. J. Rodger, N. R. Thomson, J. B. Brundell, T. Ulich, J. Lichtenberger, N. Cobbett, A. B. Collier, F. W. Menk, A. Seppälä, P. T. Verronen, and E. Turunen (2009), Remote sensing space weather events: Antarctic-Arctic Radiation-belt (Dynamic) Deposition-VLF Atmospheric Research Konsortium network, *Space Weather*, *7*, S04001, doi:10.1029/2008SW000412.

Collier, A. B. (2009), Principal component analysis of sub-ionospheric propagation conditions, *IET 11th International Conference on Ionospheric Radio Systems and Techniques (IRST 2009)*, pp. 86–90, doi:10.1049/cp.2009.0041.

Cover, T. M., and P. E. Hart (1967), Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, *13*(1), 21–27, doi:10.1109/TIT.1967.1053964.

Han, F., and S. A. Cummer (2010a), Midlatitude nighttime D region ionosphere variability on hourly to monthly time scales, *Journal of Geophysical Research: Space Physics*, *115*, A09323, doi:10.1029/2010JA015437.

Han, F., and S. A. Cummer (2010b), Midlatitude daytime D region ionosphere variations measured from radio atmospherics, *Journal of Geophysical Research: Space Physics*, *115*, A10314, doi:10.1029/2010JA015715.

Kaiser, H. F. (1960), The application of electronic computers to factor analysis, *Educational and Psychological Measurement*, *20*(1), 141–151, doi:10.1177/001316446002000116.

Lynne, K. J. W. (2010), VLF waveguide propagation: The basics, in *Propagation Effects of Very Low Frequency Radio Waves: Proceedings of the 1st International Conference on Science with Very Low Frequency Radio Waves: Theory and Observations*, edited by S. K. Chakrabati, pp. 3–41, AIP Conference Proceedings, doi:10.1063/1.3512893.

Rodger, C. J., M. A. Clilverd, N. R. Thomson, R. J. Gamble, A. Seppälä, E. Turunen, N. P. Meredith, M. Parrot, J. A. Sauvaud, and J.-J. Berthelier (2007), Radiation belt electron precipitation into the atmosphere: Recovery from a geomagnetic storm, *Journal of Geophysical Research*, *112*, A11307, doi:10.1029/2007JA012383.

Rodger, C. J., M. A. Clilverd, A. J. Kavanagh, C. E. J. Watt, P. T. Verronen, and T. Raita (2012), Contrasting the responses of three different ground-based instruments to energetic electron precipitation, *Radio Science*, *47*, RS2021, doi: 10.1029/2011RS004971.

Shao, X.-M., E. H. Lay, and A. R. Jacobson (2013), Reduction of electron density in the night-time lower ionosphere in response to a thunderstorm, *Nature Geoscience*, *6*(1), 29–33, doi:10.1038/NGEO1668.

Simon Wedlund, M., M. A. Clilverd, C. J. Rodger, K. Cresswell-Moorcock, N. Cobbett, P. Breen, D. Danskin, E. Spanswick, and J. V. Rodriguez (2014), A statistical approach to determining energetic outer radiation-belt electron precipitation fluxes, *Journal of Geophysical Research*, *119*, A019715, doi:10.1002/2013JA019715.

Thomson, N. R. (1993), Experimental daytime VLF ionospheric parameters, *Journal of Atmospheric and Solar-Terrestrial Physics*, *55*(2), 173–184, doi:10.1016/0021-9169(93)90122-F.

Thomson, N. R., M. A. Clilverd, and W. M. McRae (2007), Nighttime ionospheric D-region parameters from VLF phase and amplitude, *Journal of Geophysical Research*, *112*, A07304, doi:10.1029/2007JA012271.

Wait, J. R. (1996a), *Electromagnetic Waves in Stratified Media*, chap. 9: V.L.F. Propagation–Theory and Experiment, reissued ed., IEEE / Oxford Univ. Press, New Jersey.

Figure 1. (a) The great circle paths of the AARDDVARK observations analyzed in this study. Green circles indicate the locations of the monitored VLF communications transmitters, with call signs indicated. Red diamonds indicate the locations of the two AARDDVARK receivers (SB - Scott Base, Antarctica and EDM - near Edmonton, Canada). (b) Observations of the NDK transmission received at the EDM antenna from October 2011 to May 2014. (c) A synthetic dataset used for analysis of the success of our QDC finding techniques. White areas of plot (b) show the place-holder values replacing unusable data. The color-scales for the two upper plots are shown to the right of each plot. (d) Data from a representative day of the synthetic dataset. The red line is the true QDC, or background, and the black line is the complete data, combining perturbations and background.

Figure 2. Magnitudes of the 2-dimensional FFT spectra for the synthetic dataset shown in Figure 1c. (a) Basic spectrum before the restrictions are applied from the FFT QDC finding technique. (b) Fully restricted spectrum. Both plots have been zoomed in to frame the central background-related spectral pattern. The color-scale is $\log_{10}$ and shown to the right of each plot.

Wait, J. R. (1996b), *Electromagnetic Waves in Stratified Media*, chap. 6: Fundamentals of Mode Theory of Wave Propagation, reissued ed., IEEE / Oxford Univ. Press, New Jersey.

Wautelet, G., and R. Warnant (2012), Local climatological modeling of ionospheric irregularities detected by GPS in the mid-latitude region, *Journal of Atmospheric and Solar-Terrestrial Physics*, *89*, 132–143, doi:10.1016/j.jastp.2012.08.015.

Figure 3. (a) Nearest neighbour distances between rows of the synthetic perturbation matrix versus the distances between the corresponding rows of the full synthetic dataset, perturbations and background combined. Vertical dashed line indicates top 10 % of full data distances. (b) Daily means for the full synthetic dataset (green line and markers), smoothed dataset (red line) and background of the dataset (black line).

Figure 4. Comparison matrices, i.e., the difference between the calculated and true QDCs, for the full synthetic dataset and three QDC finding techniques. The technique used to calculate the corresponding QDC is given in the top left of each plot. All plots are on the same color-scale, which is shown to the right of the plots.

Figure 5. Comparison matrices for the smoothed synthetic dataset and subsequent application of the three QDC finding techniques. The technique used to calculate the corresponding QDC is given in the top left of each plot. All plots are on the same color-scale, which is shown to the right of the plots and is smaller than that of Figure 4.

Figure 6. (a) Synthetic data for one day (black line) and the calculated QDCs found by the smoothing process (blue line) and FFT technique (red line). (b) Perturbations in the dataset and the remainders from the techniques. (c) Comparisons between the calculated and true QDCs. All three plots have a guide bar as to the level of light on the path, either fully sunlit (light-grey), fully dark (dark-grey), or mixed with the terminator located across the path (mid-grey). The date of the day is given in the x-axis label to allow cross-checking with Figure 1c.

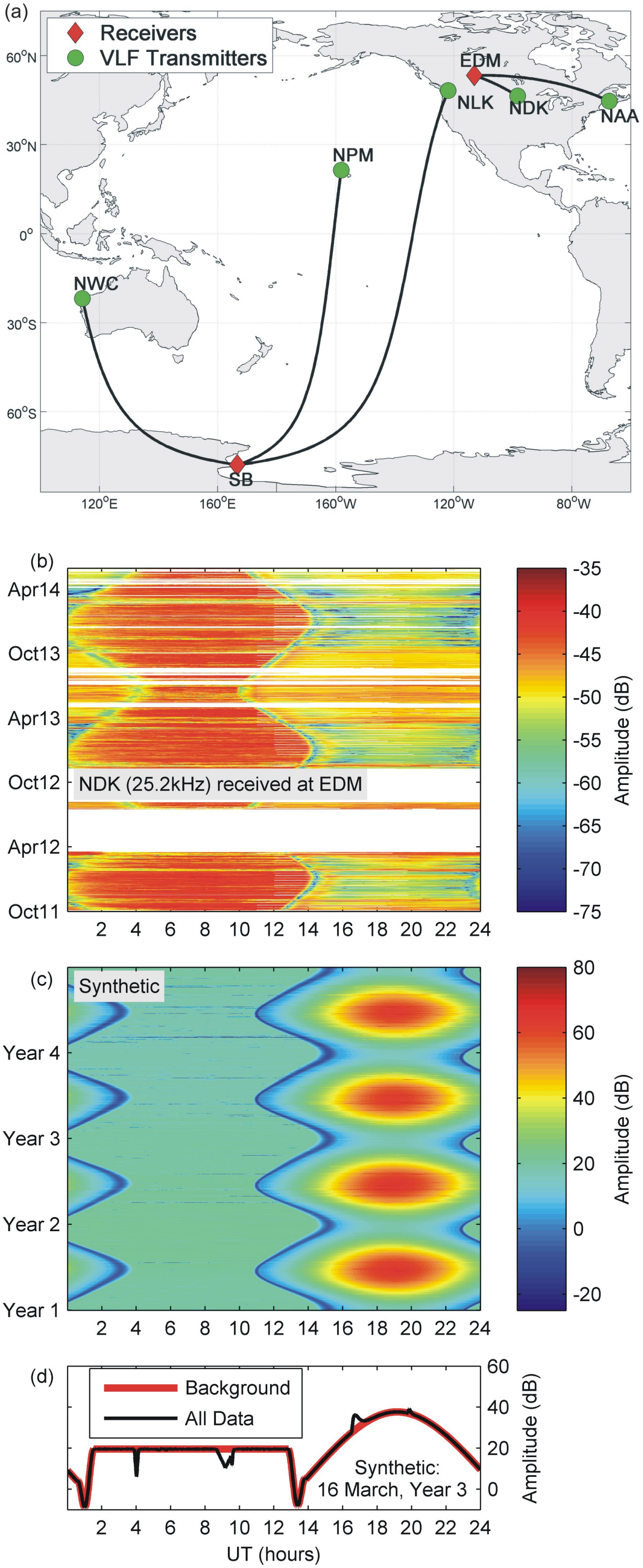
Figure 7. (a) Observations of the NWC transmission received at the SB antenna from January 2009 to December 2013. White areas of the plot show the place-holder values replacing unusable data. (b) The QDC calculated using the pre-smoothed FFT technique. The color-scale for the dataset and QDC plots is given to the right of the QDC plot. (c) The remainder, or difference between the dataset and the calculated QDC, with color-scale to the right of the plot.

Figure 8. (a) Remainders (observed amplitudes - calculated QDC) for three transmitter signals observed by the Scott Base receiver (solid colored lines, left y-axis) for 17–24 UT on 19 January 2010. (b) Remainders for two transmitter signals observed by the Edmonton receiver for 17–24 UT on 5 November 2011. Included on the plot are solar X-ray observations (thick dashed black line, right y-axis) from (a) the GOES-14 satellite and (b) the GOES-15 satellite. Grey dashed horizontal line indicates 0 dB remainder, i.e., where the calculated QDC equals the data. Grey dashed vertical lines indicate the peak flux times for NOAA identified solar flares, with the magnitude of each flare given at the base of each line.

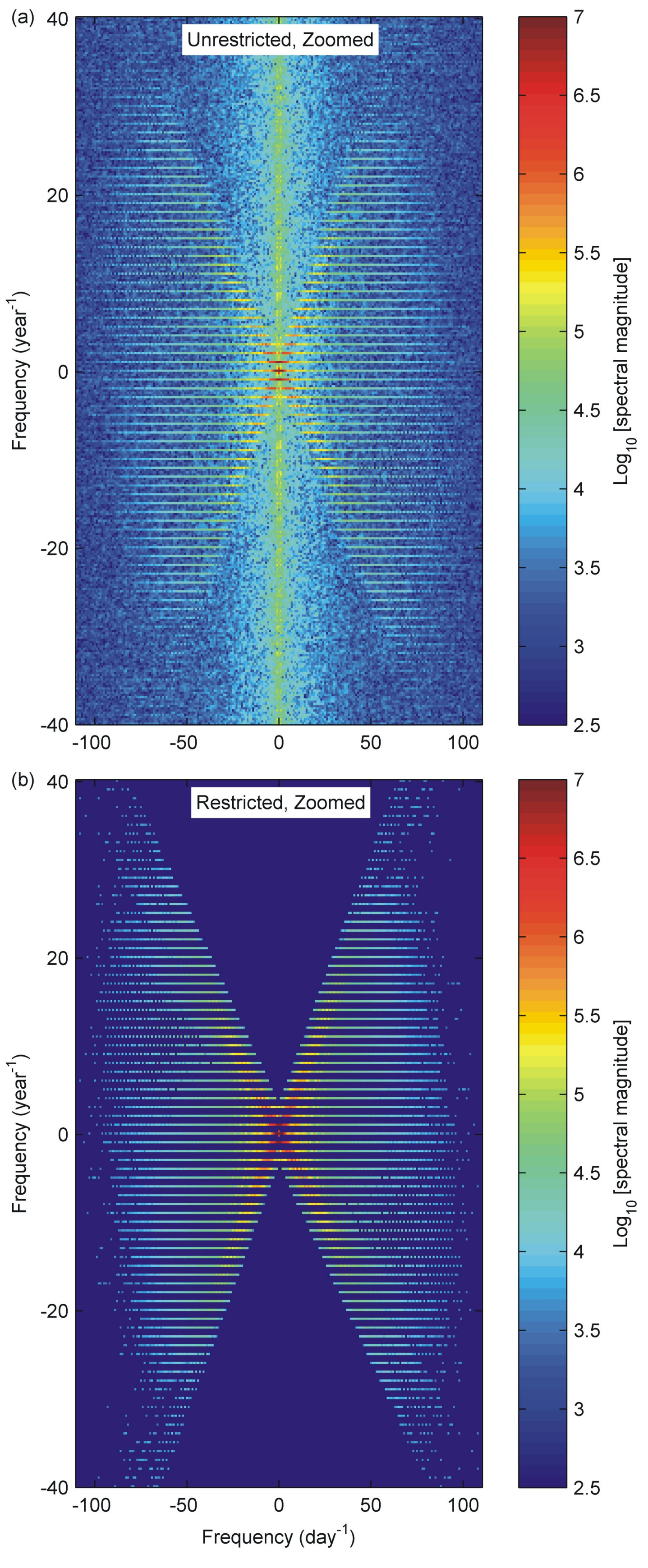
Table 1. Norms (Equation 4.1) for the comparison of our calculated QDCs to the true QDCs of our synthetic datasets. All values are rounded to 2 decimal points. Units are dB.

	All	Day	Night	Twilight
Data	2.65 ± .08	1.71 ± .04	3.45 ± .14	1.74 ± .15
CDC	1.48 ± .03	1.24 ± .01	1.33 ± .08	2.86 ± .01
PCA	2.53 ± .08	1.59 ± .04	3.24 ± .14	2.36 ± .10
FFT	0.85 ± .01	0.82 ± .01	0.9 ± .02	0.68 ± .01
	All	Day	Night	Twilight
Smoothed	0.57 ± .03	0.44 ± .05	0.24 ± .02	1.61 ± .14
CDC	0.96 ± .02	0.64 ± .02	0.29 ± .01	2.95 ± .05
PCA	0.79 ± .04	0.54 ± .03	0.47 ± .02	2.17 ± .12
FFT	0.28 ± .01	0.23 ± .02	0.14 ± .01	0.77 ± .05

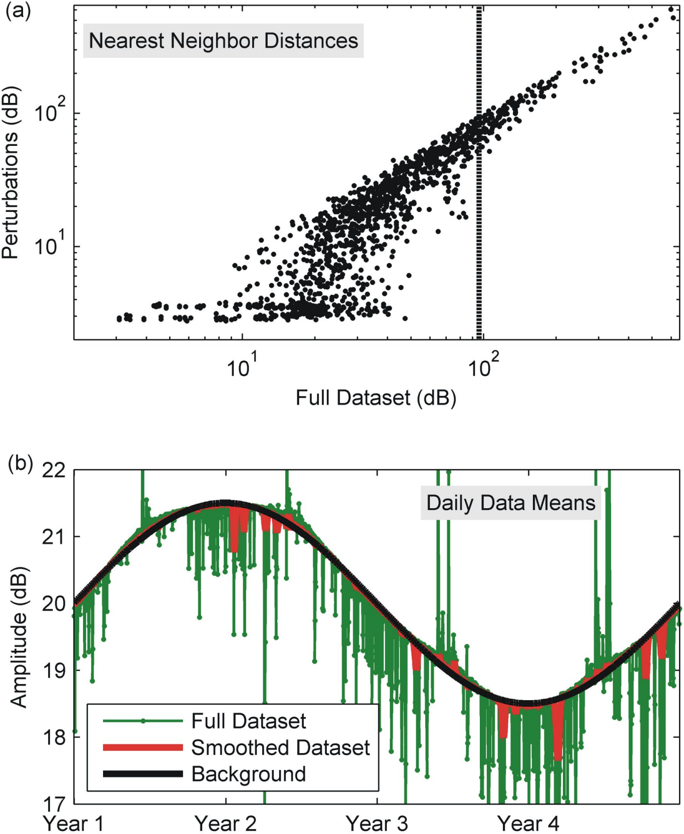
Figure 9. (a) Data from the NLK transmission observed by the Scott Base receiver (black line), and the calculated QDC (red line) for the period of a SPE starting 26 November 2011. The background color indicates the level of light on the path, either fully sunlit (light-grey), or with the terminator located across the path (mid-grey). (b) Remainder during periods when the path is fully Sun-lit or mostly dark, with the background color indicating the light level. (c) Corrected >10 MeV Proton flux observations from GOES-13. The y-axis of this plot is a $\log_{10}$ scale. The threshold for SPE recognition is marked by a horizontal dashed black line. In all plots the dashed vertical blue line indicates the time of onset of the initial flux increase, the green line the time when the SPE threshold was exceeded, and the red line the time of peak proton flux.



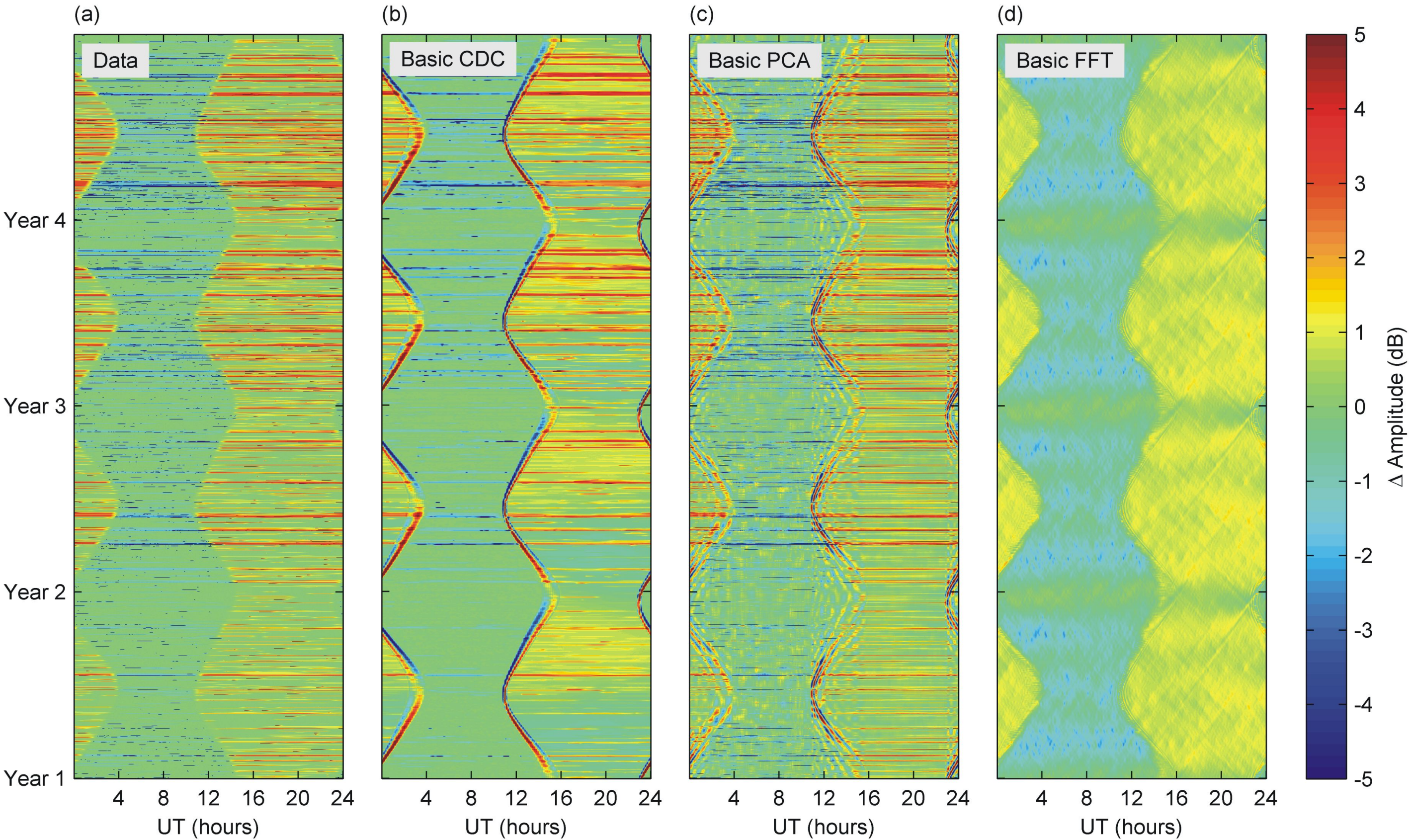
(a)
Receivers
VLF Transmitters
EDM
NLK
NDK
NAA
NPM
NWC
SB
60°N
30°N
0°
30°S
60°S
120°E
160°E
160°W
120°W
80°W
(b)
Apr14
Oct13
Apr13
Oct12
Apr12
Oct11
NDK (25.2kHz) received at EDM
Amplitude (dB)
(c)
Synthetic
Year 4
Year 3
Year 2
Year 1
Amplitude (dB)
(d)
Background
All Data
Synthetic:
16 March, Year 3
Amplitude (dB)
UT (hours)

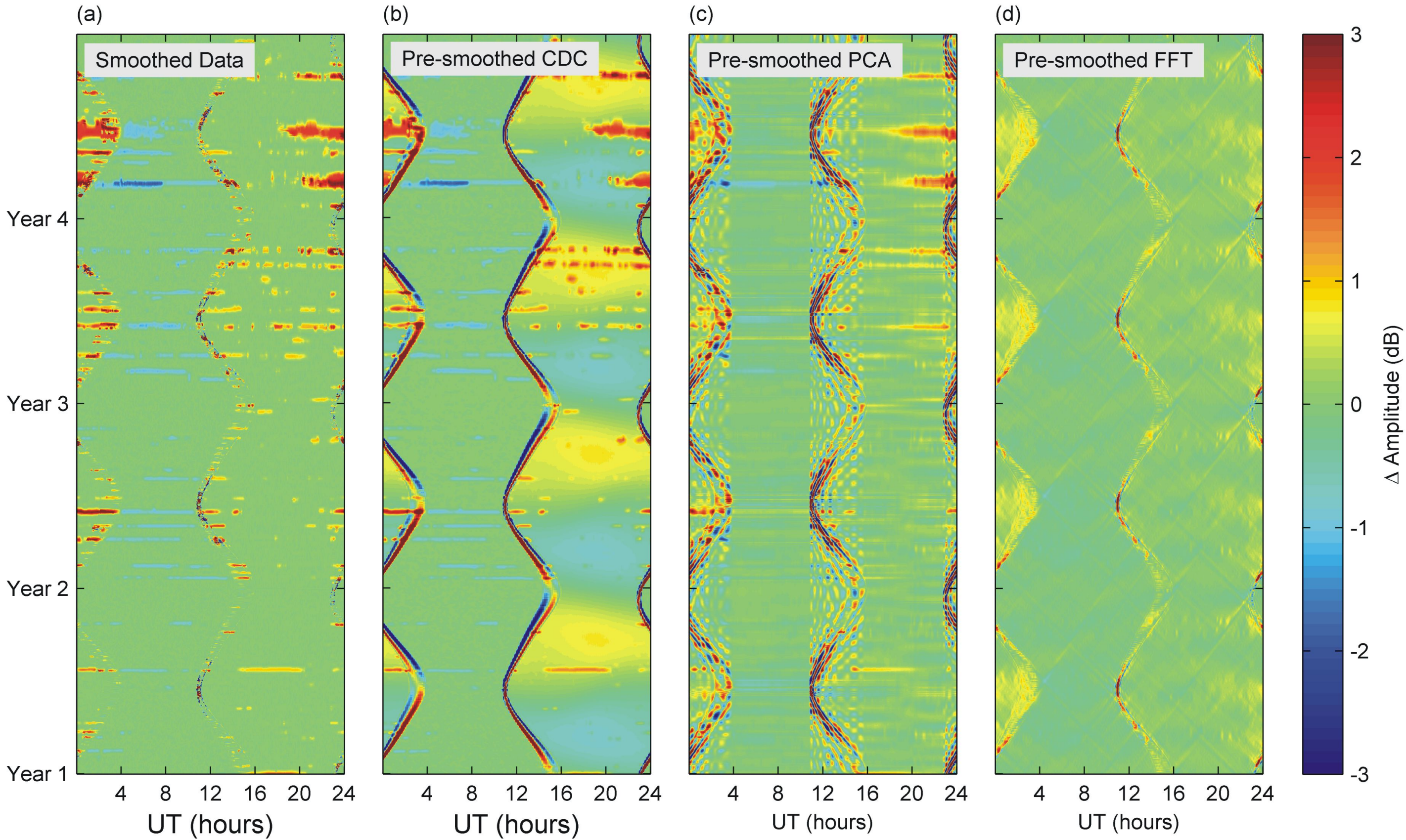


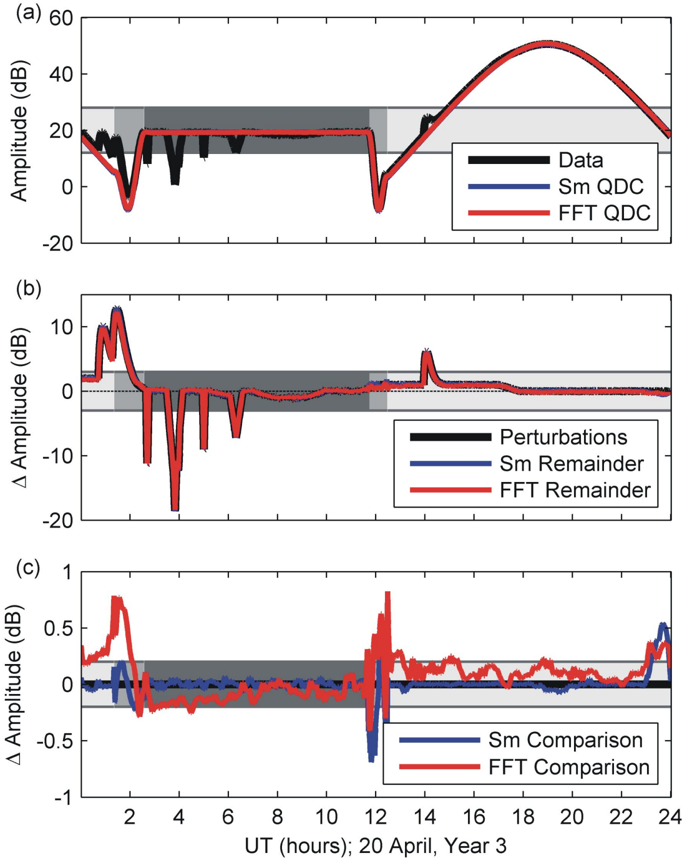
(a)
Unrestricted, Zoomed
Frequency (year^{-1})
Log$_{10}$ [spectral magnitude]
40
20
0
-20
-40
-100
-50
0
50
100
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
(b)
Restricted, Zoomed
Frequency (year^{-1})
Log$_{10}$ [spectral magnitude]
Frequency (day^{-1})

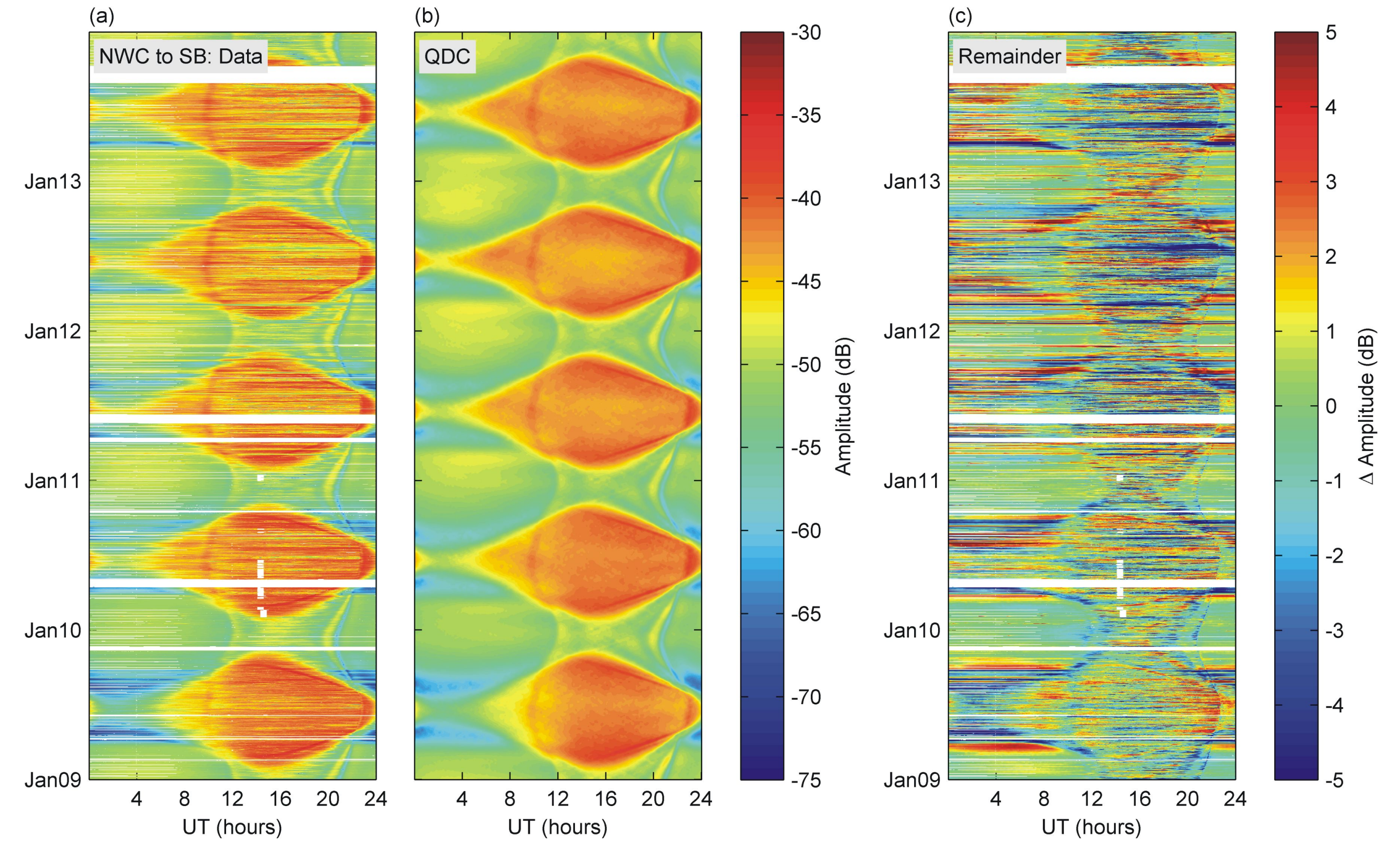


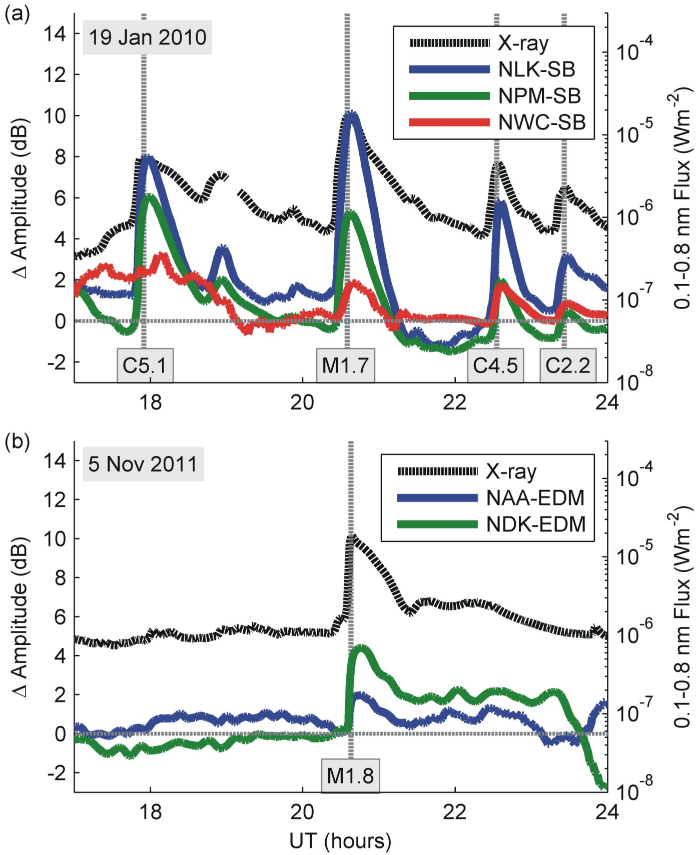
(a)
Nearest Neighbor Distances
Perturbations (dB)
10^2
10^1
10^1
10^2
Full Dataset (dB)
(b)
22
21
20
19
18
17
Daily Data Means
Amplitude (dB)
Full Dataset
Smoothed Dataset
Background
Year 1
Year 2
Year 3
Year 4

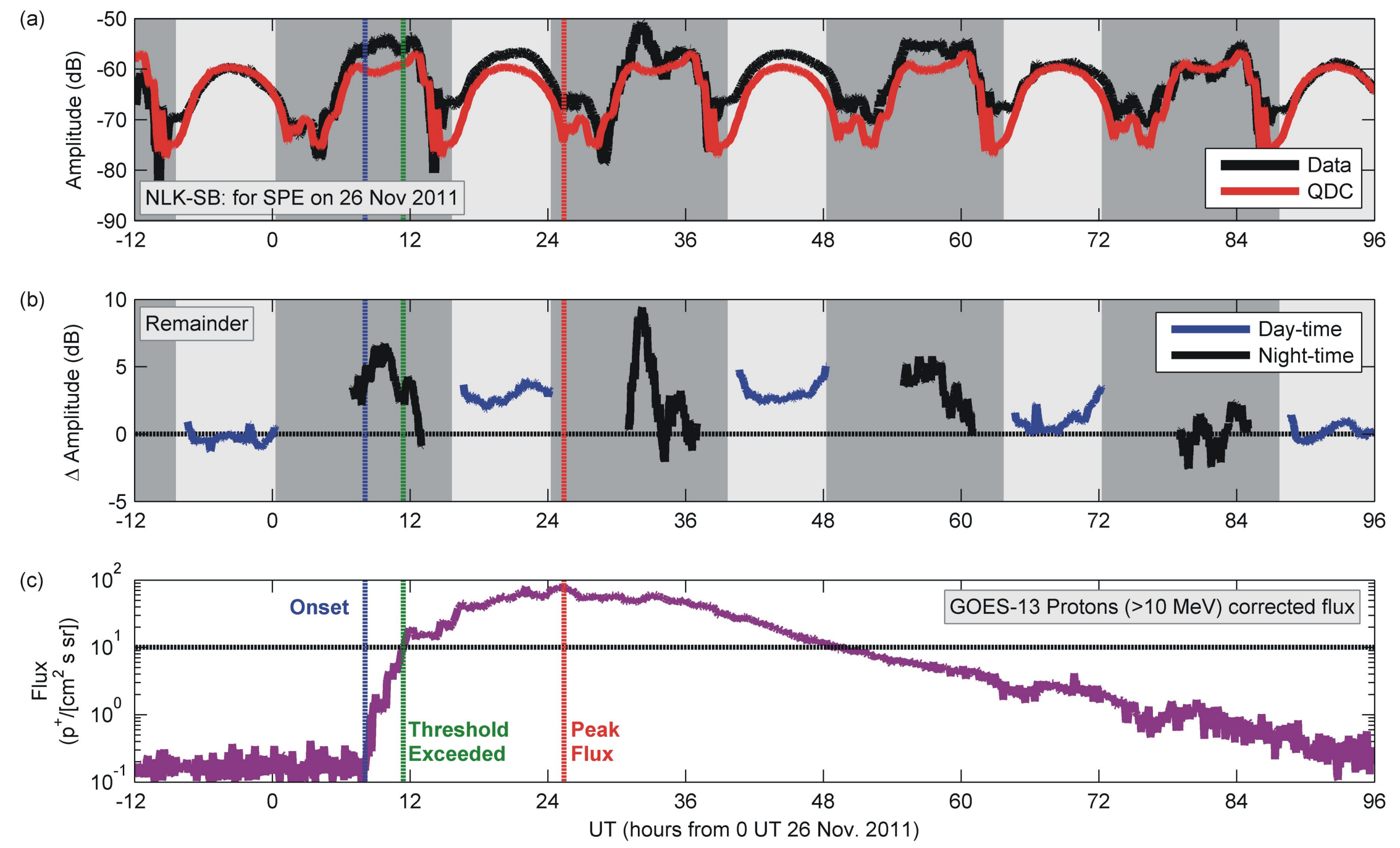

(a)
(b)
(c)
(d)
Data
Basic CDC
Basic PCA
Basic FFT
Year 1
Year 2
Year 3
Year 4
4
8
12
16
20
24
UT (hours)
Δ Amplitude (dB)
5
4
3
2
1
0
-1
-2
-3
-4
-5


(a)
Smoothed Data
(b)
Pre-smoothed CDC
(c)
Pre-smoothed PCA
(d)
Pre-smoothed FFT
Year 1
Year 2
Year 3
Year 4
4
8
12
16
20
24
UT (hours)
Δ Amplitude (dB)
3
2
1
0
-1
-2
-3


(a)
Amplitude (dB)
Data
Sm QDC
FFT QDC
(b)
Δ Amplitude (dB)
Perturbations
Sm Remainder
FFT Remainder
(c)
Δ Amplitude (dB)
Sm Comparison
FFT Comparison
UT (hours); 20 April, Year 3


(a)
NWC to SB: Data
(b)
QDC
(c)
Remainder
Jan13
Jan12
Jan11
Jan10
Jan09
4
8
12
16
20
24
UT (hours)
-30
-35
-40
-45
-50
-55
-60
-65
-70
-75
Amplitude (dB)
5
4
3
2
1
0
-1
-2
-3
-4
-5
Δ Amplitude (dB)


(a)
19 Jan 2010
X-ray
NLK-SB
NPM-SB
NWC-SB
C5.1
M1.7
C4.5
C2.2
Δ Amplitude (dB)
0.1-0.8 nm Flux (Wm^{-2})
(b)
5 Nov 2011
X-ray
NAA-EDM
NDK-EDM
M1.8
UT (hours)



(a)
Amplitude (dB)
NLK-SB: for SPE on 26 Nov 2011
Data
QDC
(b)
Δ Amplitude (dB)
Remainder
Day-time
Night-time
(c)
Flux
(p+/[cm2 s sr])
Onset
Threshold
Exceeded
Peak
Flux
GOES-13 Protons (>10 MeV) corrected flux
UT (hours from 0 UT 26 Nov. 2011)