A METHOD FOR THE AUTOMATIC SELECTION OF SOLAR TYPE III RADIO BURSTS WITH WIND/WAVES

J. E. Waters[1*], C. M. Jackman[2], A. R. Fogg[2], C. K. Louis[2], L. Lamy[3,1], C. Briand[3], X. Bonnin[3], M. Maksimovic[3], P. C. Murphy[3], B. Cecconi[3], K. Issautier[3], and D. K. Whiter[4]

*Corresponding author: james.waters@lam.fr

Citation:
Waters et al., 2023, Automatic Selection of Solar Type III Radio Bursts with Wind/WAVES, *in Planetary, Solar and Heliospheric Radio Emissions IX, edited by C. K. Louis, C. M. Jackman, G. Fischer, A. H. Sulaiman, P. Zucca, published by DIAS, TCD*, pp. 85–103, doi: 10.25546/104053

Abstract

Solar type III radio bursts have a characteristic signature in frequency-time dynamic spectrograms and provide important insight into the dynamics of the Sun and solar wind. Direct physical inferences of the source electrons or their propagation can be made, but radio observations typically contain multiple emission sources which must be accounted for. As well as this, the emission feature of interest must be selected, which can be problematic if done manually, particularly with a large amount of data. The WAVES instrument onboard the Wind spacecraft has accumulated almost thirty years of solar and terrestrial observations, with the majority from L1. Our previous work (Waters et al., 2021b) has focused on the extraction of terrestrial auroral kilometric radiation (AKR) from the Wind/WAVES data, and the removal of noise. In this study we use a similar concept, separating the radio sources by the measured signal variability during a spacecraft spin, but focusing on the parameter space containing interplanetary (IP) type III bursts below 1 MHz. Using an inverse threshold of the spin variability metric and a signal-to-noise ratio (SNR) threshold, the radio data is filtered to return mostly Type III emission above 200 kHz. Integrating the flux density between 500-1045 kHz and choosing a threshold of 1×10^9 W sr^{-1} to select start times of Type III bursts between 1995 and 2021

[1] *Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France*
[2] *School of Cosmic Physics, DIAS Dunsink Observatory, Dublin Institute for Advanced Studies, Dublin 15, Ireland*
[3] *Observatoire de Paris, LESIA, PSL Research University, CNRS, Sorbonne Université, University of Paris, Meudon, France*
[4] *Department of Physics and Astronomy, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK*

yields the detection of 120560 events. Annual event occurrence shows good correlation with yearly-averaged sunspot number across the two solar cycles examined here (P=0.94). Evaluating classification statistics of the selection via visual examination of a random sample of 50 days of Wind data returns a true positive rate of 0.9 and a positive predictive value of 0.4, indicating that the majority of observed Type III bursts are returned. This technique thus provides an initial step to broad statistical examination of the physical properties of Type III bursts using observations from Wind/WAVES or other instruments.

1 Introduction

Solar Type III radio bursts are generated from electron beams that are accelerated near the solar surface and travel along open interplanetary magnetic field (IMF) lines (Wild, 1950; Dulk, 2000). In frequency-time dynamic spectrograms, when considering kilometric wavelengths ($\sim$ 10 - 1000 kHz), Type III bursts appear as a bright, transient feature first at higher frequencies ($\sim$ 1 MHz), that then rapidly reaches lower frequencies (between $\sim$ 50-100 kHz) where longer drifting features are often seen for shock-accelerated, IP Type III bursts that reach Earth (e.g., Wild et al., 1963; Stewart, 1974; Hoang et al., 1995; Dulk, 2000). In this work, we refer exclusively to IP Type III bursts when discussing the Wind observations. This characteristic and easily visually recognisable signature represents the path of the electron beam; Type III emission indicates the presence of an active region on the solar surface and analysis of the spectral information can provide important physical insights to the source electron stream and interplanetary plasma (Dulk, 2000). Energetic Type III bursts are related to solar flares, and as such become more frequent and intense when occurring near solar maximum, and becoming more intense with solar energetic particle events (Cliver & Ling, 2009).

For the ongoing study of Type III bursts, it is ideal to have as complete a list as possible of their occurrences based on available observing instruments. With such a list, the spectral and temporal properties of Type III bursts can be explored statistically. Type III bursts can appear in observations as single, distinct emissions, in groups of a few with varying intensity, or in a quasi-continuous manner during Type III storms, where thousands of weak Type III bursts can appear in a day (Potter et al., 1980; Reiner et al., 2007). Properties of all classes of Type III bursts have been explored in the past; in investigations of the decay time of individual bursts (Krupar et al., 2018), or with properties of Type III storms such as waiting time and brightness (Eastwood et al., 2010) for example. With a proper characterisation of the spectral information of the Type III and access to the appropriate instrumentation, goniopolarimetric inversions can be applied and subsequently return important parameters such as the polarisation (Hanasz et al., 1980; Cecconi & Zarka, 2005; Reiner et al., 2007; Cecconi et al., 2008). Directivity effects can also be characterised by taking advantage of the range of observing positions with space-based instrumentation (Bonnin et al., 2008).

The Wind spacecraft has provided more than 20 years of data to allow monitoring of the solar wind plasma and other solar emissions since beginning operation at the start of

1995, and has been following halo orbits about the Lagrangian point L1 since early 2004. Equipped with the WAVES radio instrument (Bougeret et al., 1995), Wind is also able to provide observations of solar Type III bursts which, since Wind reached L1, are typically comprehensive due to the usual intensity of the Type IIIs and the relative lack of other radio sources to contaminate the observations made from this region. Based on a recent tool that has been developed to extract observations of terrestrial AKR from the entire Wind/WAVES data archive (Waters et al., 2021b,a; Fogg et al., 2021, 2022), we introduce a simple selection technique, based on an inverse binary mask for the AKR selection, that can be applied to Wind/WAVES data and expedite the cataloguing and study of Type III bursts.

While the automatic selection of Type III bursts to facilitate statistical analysis has been explored previously with the use of techniques such as the Hough transform and other methods (Lobzin et al., 2014; Zhang et al., 2018; Singh et al., 2019; Scully et al., 2023), this computationally inexpensive and relatively simple technique can facilitate the retrieval of Wind Type III observations; bolstering the literature with a set of observations from a consistent viewing point of L1. Other catalogues typically use image processing techniques that require heavy computation, or are hindered by irregular observation windows and constraints on frequency sampling due to their location on the ground. The catalogue by Zhang et al. (2018) is one such technique; derived from observations from the Nançay Decametric Array (NDA), which performs high resolution measurements of solar radio bursts between 10-80 MHz when directed towards the Sun, and covers the period between 2012-2017. Such a catalogue provides an invaluable resource for both the study of such bursts but also as a comparative measure for event lists such as this one.

To illustrate the technique, we have applied the selection to Wind data between 1995 to 2021, covering two solar cycles. In Section 2, we describe the instrumentation, dataset, and conceptual steps of the selection process. In Section 3 we present some exemplary output of the technique and evaluate the selection performance; 3.2 describes the event occurrence and calculates binary classification performance metrics, while 3.3 compares the temporal discrepancy and event occurrence between corresponding bursts in this catalogue and the work of Zhang et al. (2018). This is followed by a discussion of future avenues for this method in Section 4.

2 Method

Wind is a spin-orbit stabilised spacecraft equipped with the radio instrument, WAVES (Bougeret et al., 1995), that itself includes 3 radio antennae and a series of receivers. Two of the WAVES antennae (E_x and E_y) are in the spin plane of Wind, while one (E_z) is aligned with the spin axis. The WAVES RAD1 receiver most often operates by observing 32 distinct frequency channels between 20-1040 kHz, each measured during a 3 second spin period of Wind, and frequently observes the lowest frequency drifting component of solar Type III radio bursts. The RAD1 receiver is the focus of discussion in this study. Figure 1 shows the orbit of the Wind spacecraft between 2000 and 2004, with panel a annotated to show points of interest. Particularly of interest is the path of Wind projected into the ecliptic plane in Figure 1a, with the Sun and the Lagrangian point L1 shown on to the

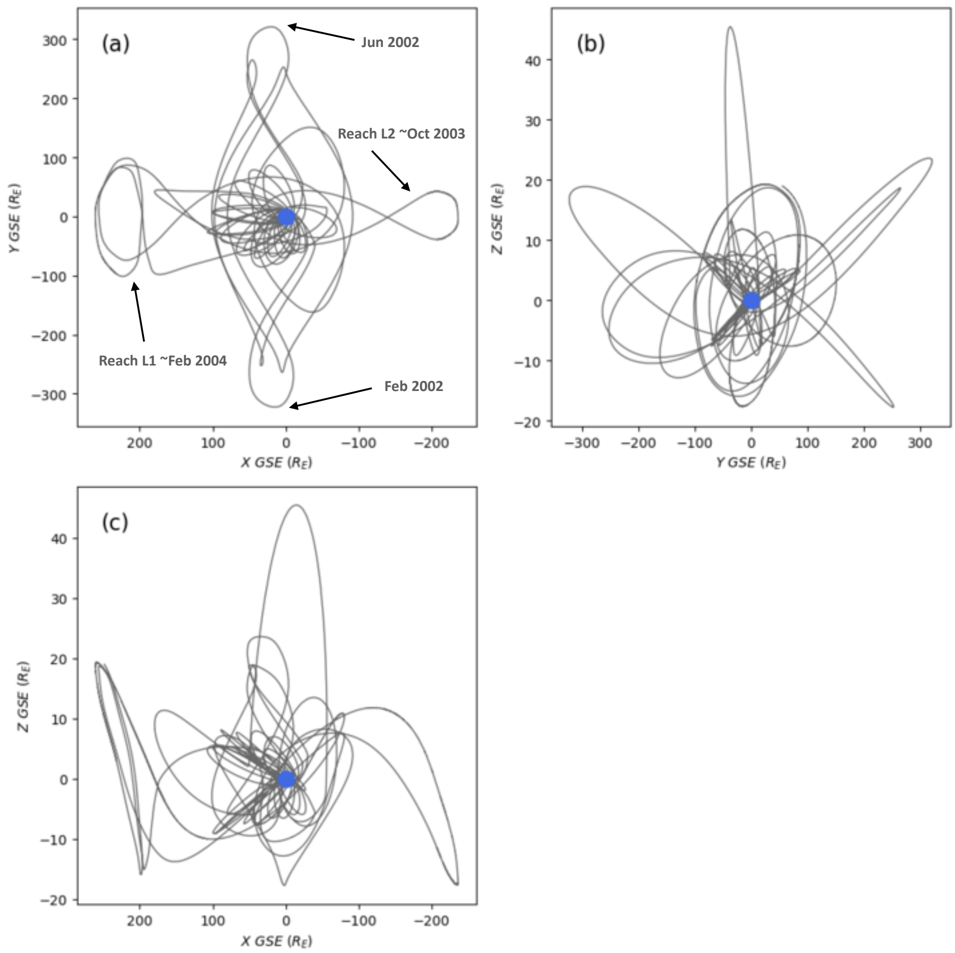

Figure 1: *Wind's orbital trajectories in GSE coordinates from 2000-2004. Panel a shows the projection to the ecliptic plane and is annotated to show dates of interest. Panels b and c show a view along the Earth-Sun line, away from the Sun, and the projection onto the noon-midnight meridian, towards dawn, respectively. Panels a and c have the Sun on the left. Panel a is annotated to show points of interest including L1. While orbits of varying eccentricities and precessing apogees were completed between Wind's launch in November 1994 and 2000 (not shown), Wind has continued to orbit L1 since February 2004 in a similar way as represented in the figure.*

left, the Lagrangian point L2 to the right, and the Earth in the centre of the figure. Since February 2004, Wind has remained in a semi-periodic halo-like orbit about L1 (Pelton & Allahdadi, 2015).

Prior to February 2004 in the figure, and in other orbits prior to 2000, Wind has explored many important regions local to Earth's magnetosphere using orbits with apogees that precess in longitude, as well as conducting extreme diversions to the deep magnetotail or solar wind as shown. Since WAVES measures the most intense radio emission incident upon the antennae, terrestrial auroral kilometric radiation (AKR) is often observed when Wind is on the nightside of Earth. AKR, the most intense terrestrial radio emission, is typically observed as bright and burst-like between frequencies of $\sim$ 40-800 kHz. The confined AKR source locations and anisotropic beaming due to the cyclotron-maser-instability generation mechanism mean that AKR is best observed from a range of local times that correspond approximately to the nightside of Earth (Fogg et al., 2022; Waters et al., 2022). From such positions, Wind tends to observe AKR primarily, with some Type IIIs. Exemplary AKR observations can be seen in Figure 3a, with the most intense burst occurring between 18:00-21:00 UT, or Figure 2 of Waters et al. (2021b). By contrast, on the dayside, the AKR is much more transient and the intensity is attenuated, and Type III bursts are observed with greater clarity.

A technique has recently been developed by Waters et al. (2021b) that allows for AKR to be automatically extracted from observations from the Wind/WAVES RAD1 receiver. With the spin axis aligned antenna only, the measurements are calibrated using a simplified form of a goniopolarimetric inversion, bespoke to AKR, and selected at each frequency based on the variability of measurements made during a spin (Waters et al., 2021a). This variability is determined using a metric known as σ_Z, which is calculated for each frequency and from each sample in a 3 minute sweep. The received power of the spin-axis aligned antenna are normalised by the spin average, and σ_Z is given by the standard deviation of these values. This technique is effective at isolating AKR observations due to the greater fluctuation of received power at the antenna, attributable to the frequent ignition of many point sources in the inner magnetosphere. It is possible that rapidly varying, short lasting solar radio emissions, whether associated with Type IIIs or not, are also selected here. Statistical results of the LT dependence of the Wind AKR suggest the selection is robust (Waters et al., 2021b; Fogg et al., 2021; Waters et al., 2022), however it must be noted that such short solar radio emissions could be removed with the method presented here.

The disparity between σ_Z values of Type III bursts and AKR indicates that the former can be characterised by a lower variability during the 3 second spin, likely due to the more consistent illumination of the spacecraft by emission from a single radio source (i.e generated from a given electron beam, travelling along an interplanetary magnetic field line) instead of sporadic illumination by many intense sources (in the case of multiple active AKR source regions). Propagation effects are also likely to play a role (Musset et al., 2021). Due to this discrepancy, and the agnosticity of the variability metric to the radio source, a mask can be applied with the inverse threshold to leave only Type III bursts and background emission. As the background spectrum is measured for each sweep of the receiver, the measured spin signal-to-noise ratio (SNR) can also be used as a criteria to further remove background emission, at all frequencies but primarily those $\gtrsim$ 200 kHz. By then setting a threshold on the integrated power, the initial time of Type III bursts can be automatically selected. In this way, a catalogue of Type III occurrences could be obtained with scope for almost real-time application to future observations, and to provide the initial event time with a maximised success rate.

3 Type III selection performance

3.1 Exemplary applications of Type III selection

Figures 2 and 3 show examples of how a combination of the inverse AKR mask and SNR thresholding can be applied to Wind observations to retrieve information regarding Type

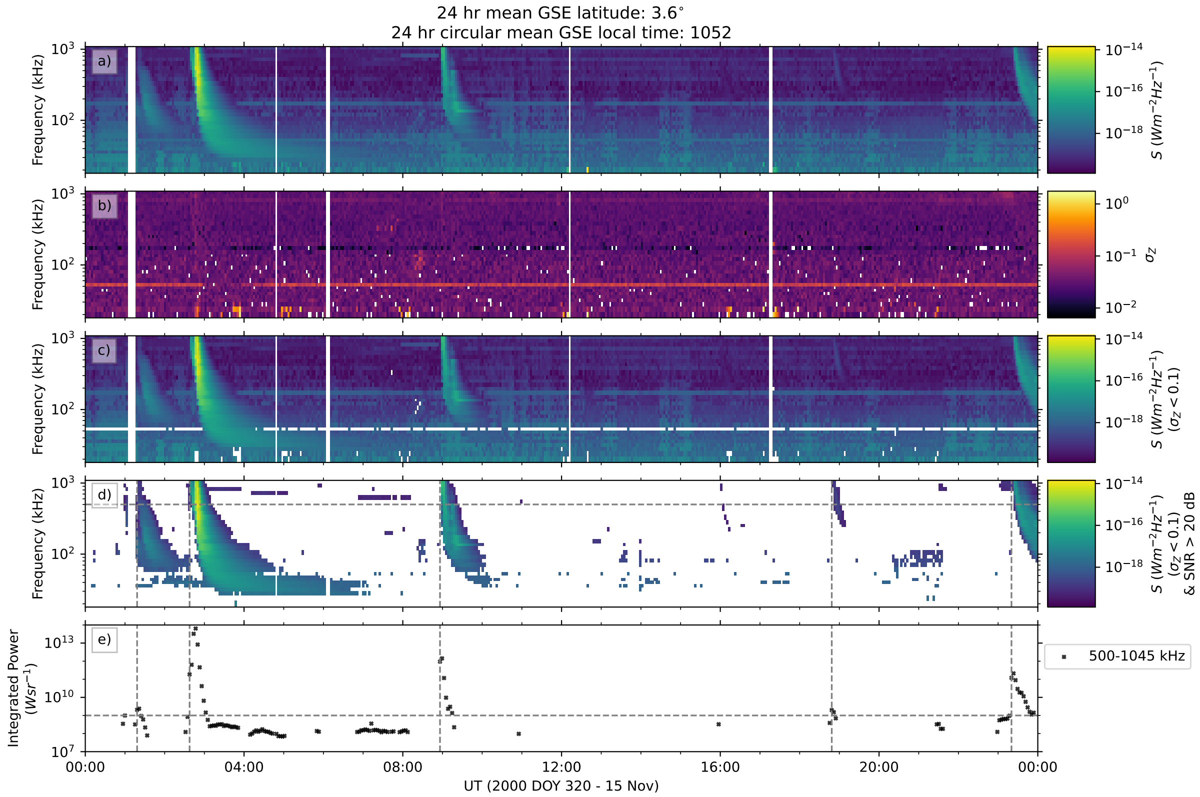

Figure 2: *Frequency-time dynamic spectrograms and derived data for 15 November 2000 that present the automatic selection of Type III emission from Wind/WAVES data. Panel (a) displays all WAVES observations, giving the flux density as calibrated by the description in Waters et al. (2021b) and so invalid for the Type III bursts. Panel (b) shows the average σ_z spectra for the period, which is the dimensionless variability metric used to differentiate AKR from other emission. Radio-frequency interference and local plasma waves are responsible for the "brightest" features in this example. Panel (c) shows the result of applying a binary mask to the flux densities using the criteria $\sigma_z < 0.1$. Panel (d) shows the result of applying a subsequent threshold on the SNR of the observations, namely SNR > 20 dB. The grey line corresponds to the lower frequency bound of the 500-1045 kHz range used to calculate the integrated power of any remaining emission, shown in panel (e). Panel (e) also shows the threshold limit used here ($P = 1 \times 10^9$ W sr^{-1}) as well as the times returned by this technique.*

III burst occurrence. Each demonstrate typical Wind/WAVES/RAD1 observations from varying spacecraft positions to highlight the efficacy of removing contaminating AKR and background emission. Figure 2 shows 24 hours of observations made when Wind was at around 1100 hours local time (LT) and near the ecliptic plane. Figure 2a shows average, AKR-calibrated flux density spectra for each 3 minute receiver sweep. Although the flux density cannot be used for physical interpretation of solar Type III bursts, the primary aim here is to select the temporal information and so the flux densities are presented in this way for convenience. It would also be possible to apply the Type III selection to the spectral power density measured at receiver level, from which point a specific calibration could be applied if required by further analyses. Figure 2b shows the selection metric σ_z that is based on the measured variability during a spin, designed to extract AKR observations by Waters et al. (2021b) (an example with the clear presence of AKR in σ_z is given in Figure 3b).

In Figure 2, no AKR is observed and therefore little emission is removed; those that are removed are done so due to the presence of radio frequency interference (RFI) or quasi-thermal noise. Figure 2c shows the result of applying a binary mask that follows the inverse threshold on the σ_z variability metric in Figure 2b. In the example, $\sigma_z < 0.1$ limit is used as a threshold; $\sigma_z \geq 0.1$ is the limit originally used for AKR selection. Figure 2d shows the same data as in Figure 2c with a threshold of 20 dB imposed on the signal-to-noise ratio (SNR), which in turn relies on a background spectrum derived over 24 hours (see (Zarka et al., 2004; Waters et al., 2021b) for details). The value of 20 dB is chosen here as that has been used before to differentiate emission that is able to be analysed with previous goniopolarimetric inversions (Cecconi & Zarka, 2005). Figure 2e shows the integrated power of the remaining emission, between 500-1045 kHz; the lower frequency is indicated by the horizontal grey line in Figure 2d and the higher frequency is the maximum observable by the RAD1 receiver. The integrated power is calculated using the AKR-calibrated flux density and is given in units of W sr^{-1} as it is difficult to apply a single solid beaming angle of a source to AKR observed by Wind. While the magnitude is technically incorrect, of value here is the relative value of the integrated power; imposing a third threshold on this can allow for the retrieval of times for which the Type III is initially observed, when the electron beam is closer to the Sun. For the purposes of studying AKR the calibrated flux density is normalised to 1 AU using the Earth-spacecraft distance; no such normalisation is present here. It should also be noted that the 500-1045 kHz range used for the power here is conservative and could be extended to lower frequencies, such as those above $\sim$ 200 kHz where other emission seems effectively removed from visual examination, to define the Type III more clearly.

Figure 3 shows another 24 hour observation to exemplify the Type III selection process, this time from measurements made by Wind when in the pre-dawn sector, around 0500 hours LT, again near the ecliptic plane. While the panel layout is equivalent to that of Figure 2, the presence of the bright AKR is clear in panels a-c. Here, AKR is seen between $\sim$ 12:00-22:30 UT; patchy bursts are observed with the most intense occurring between $\sim$ 18:00-21:00 UT over $\sim$ 60-700 kHz. The effect of the σ_z selection metric is now seen, highlighting the AKR, which is subsequently removed with the application of the $\sigma_z < 0.1$ threshold. Using the same criteria as before, the SNR threshold applied in panel d can be used to return mostly isolated Type III observations where the brightest emission is attributable to these solar sources, while the integrated power can be used to automatically retrieve initial temporal information about Type III bursts across Wind observations.

To demonstrate the efficacy of the Type III selection described here the algorithm was applied to 27 years of Wind/WAVES data, covering two solar cycles, between 1st Jan 1995 and 31st Dec 2021. An integrated power threshold of 1×10^9 W sr^{-1} was used over the frequency range of 500-1045 kHz, and the times of initial 3 minute spectra that exceed this threshold are returned as event times. In the following subsection 3.2 the event list, occurrence statistics across the solar cycle and the classification statistics are examined, while in the subsection 3.3 the event list is compared with a list of Type III bursts inferred from routine solar observations of the NDA. We note that the threshold of 1×10^9 W sr^{-1} was chosen after evaluating the results of six logarithmically-spaced values between 5×10^8 W sr^{-1} and 1×10^{11} W sr^{-1}. Evaluation was based on a qualitative

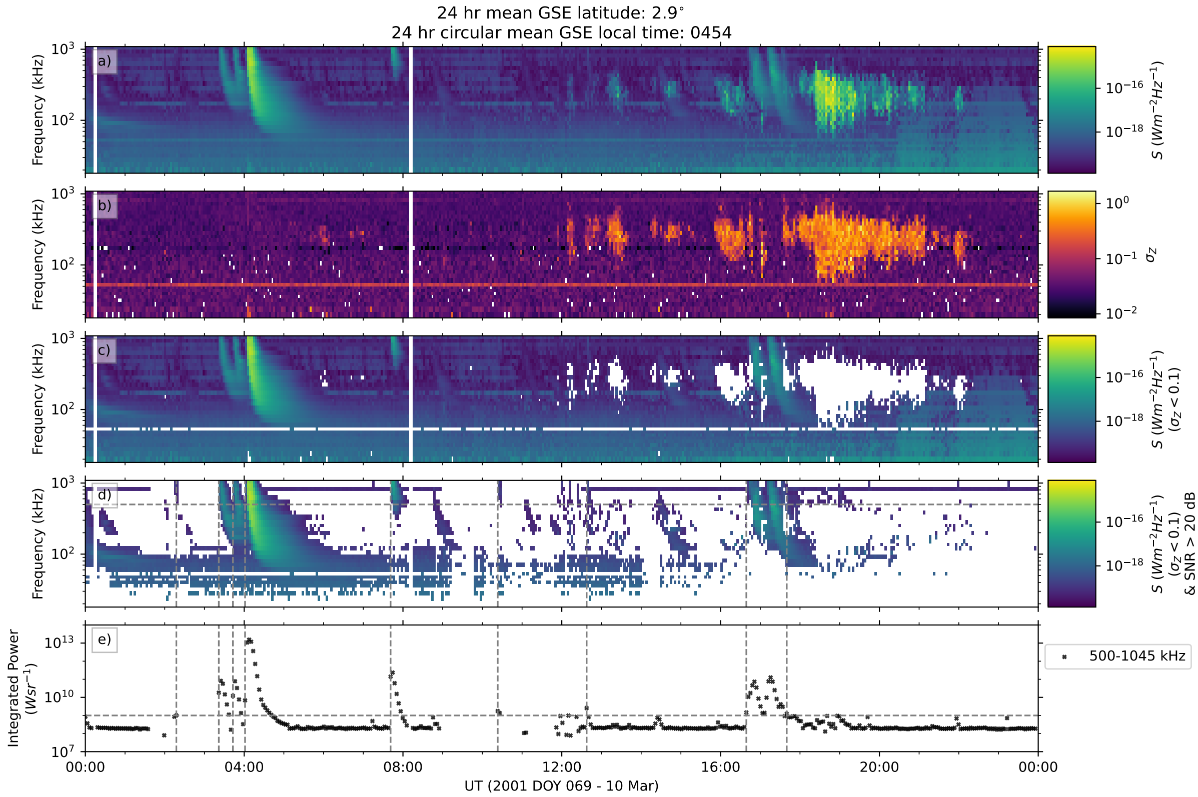

Figure 3: *Further 24 hour example of the automatic Type III selection, from 10 Mar 2001, showing significant contamination by AKR between 12:00 and 22:00 UT. Each panel presents similar data as described in Figure 2. The presence of AKR and the effect of the binary mask can clearly be seen in panels (b) and (c).*

assessment of correlation between event occurrence and the solar cycle (as in the beginning of Section 3.2) and of the temporal discrepancy with the published catalogue of Type III bursts observed with NDA (as in the beginning of Section 3.3). Higher threshold values are more restrictive and include brighter bursts; the threshold value of 1×10^9 W sr^{-1} was chosen to minimise the temporal discrepancy and produce as best a correspondence as possible, while limiting the inclusion of ambiguous emission. The results of other thresholds are not detailed here for brevity.

3.2 Event occurrence and classification statistics

With a threshold of 1×10^9 W sr^{-1}, 120560 bursts are returned from the selection algorithm. Solar Type III radio bursts have been shown to correlate with general solar activity throughout the $\sim$ 11 year solar cycle (Saint-Hilaire et al., 2013), following from their association with active regions on the solar surface. The breadth of this data thus provides the opportunity to examine the occurrence of Type III bursts over two solar cycles, as well as measure the selection performance by comparing with a proxy of solar activity. Here events are compared with the sunspot number (`https://sidc.be/SILSO/home`), as shown in Figure 4.

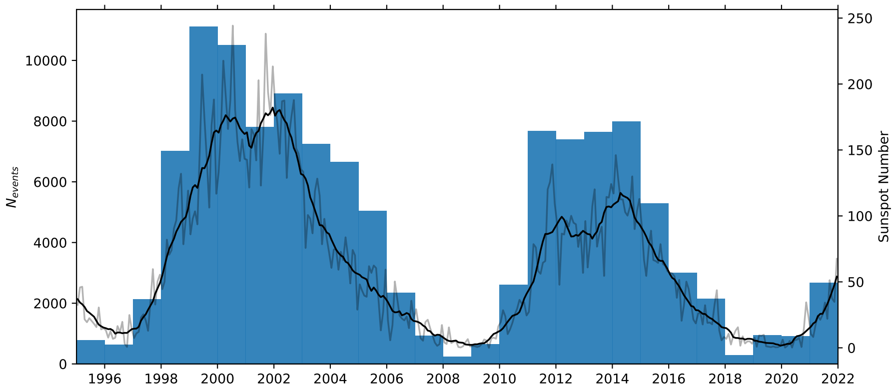

Figure 4: *Occurrence distribution of Type III bursts selected by this method between 1995 to 2021. Events are binned in 12 month intervals. On thge right-hand ordinate the monthly mean sunspot numbers, provided by the Royal Observatory of Belgium (`https://sidc.be/SILSO/home`) are plotted in grey as a proxy for the 2 solar cycles covered by this period. A rolling average of the sunspot number, taken over a 12 month period, is also plotted in black.*

There are maxima in the number of events returned by the algorithm around the solar maxima in 2000 and 2014, with fewer events at solar minima. While the event occurrence does not follow the sunspot number exactly throughout the cycle, there is good correlation. Using the yearly-averaged sunspot number (not shown) we calculate a Pearson correlation coefficient of 0.94 with the Type III occurrence. The number of Type III bursts returned from this selection also decreases with the lower solar activity of solar cycle 24, between 2008 and 2019. The temporal correlation of event occurrence with the solar cycle is an indicator of the broad trends in our catalogue.

To explore the performance of the selection in a quantitative manner it is possible to take a published list of Type III bursts to compare to, such as that curated by Zhang et al. (2018), or examine output of the event selection output manually to evaluate the presence of "correct" classifications by eye. While the latter is open to bias from subjective evaluation, and can only feasibly be performed on a subset of the classifications due to time constraints, we feel that this is more robust to quantify the classification performance in this case. To the authors knowledge there is no published list of Type III bursts that uses Wind/WAVES, where the observations would be directly comparable. Thus, a catalogue of ground-based Type III observations would need to be used as a set of "true" observations, even though discrepancies in the sampled frequencies of observations are present. These are described further in Section 3.3.

To quantify the classification of the performance of this technique, 50 dates, each covering 24 hours of observations (excluding invalid data), are chosen at random between 1995-2021. The selection of dates was examined to ensure even sampling of both solar cycles, as well as the local time position of Wind due to the wide data coverage and large number of observing positions during the 27 years. After choosing the dates, the dynamic spectrograms, such as those shown in panels a and d of Figures 2 and 3, were visually cross-compared with the burst lists. In this way it is possible to determine the number of true positives, where the selection classifies Type III bursts that have a distinct intensity and morphology, false positives, where the selection classifies emission that does not appear to be a Type III burst, has limited spectral extent or is ambiguous (when part of a Type III storm for example), and false negatives, where a Type III burst is observed but the spectra corresponding to the highest frequency initial portion of the burst is not selected.

In Figure 2 for example, all classifications would be counted as true positives. Although the bursts exhibit a range of intensities, the spectral morphology is clear. This is more ambiguous for the emission at ∼19:00 UT, but this is a good example of a limiting case for a visual examination, where the emission is noticeable in the whole spectrogram (Figure 2a) and is highlighted by the filtered data (Figure 2d). Figure 3 exhibits multiple true positives, such as those between 03:00-04:00 UT, as well as cases of false positives and true negatives. The burst at ∼08:00 UT would be counted as a true positive due to similar reasons as the limiting case discussed above, while those at ∼02:00 UT and ∼10:00 UT would be counted as false positives due to the ambiguity of the emission; they do not clearly follow a drifting structure and are not intense enough to be distinct in the complete spectrogram in Figure 3a. A true positive is also seen at ∼17:00 UT, where the first of a group of two Type III bursts is classified. The returned event that follows is a false positive (as it does not classify the start of a burst), while the second burst of the group in the spectrogram is missed by this technique and would therefore be counted as a false negative.

Although it is relatively easy to identify the presence of a Type III storm, it is sometimes difficult to perceive a distinct Type III burst as mentioned above; this contributes to the bias of such a visual examination. Despite this, false positives are often seen during Type III storms or when instrumental or local noise is present. In these cases, the integrated power between 500-1045 kHz is often above or close to the threshold, and background emission or weak Type III bursts are then returned as false positives. It could also be possible that lower frequency Type II burst emission is selected, creating a false positive. This is due to the power falling below the threshold before returning to typical values, above the threshold. Bright Type III bursts that have a higher relative power are missed if the threshold has been previously exceeded by the "background", which in general also increases the number of false negatives. This is also the case when two Type III bursts occur close together. An example of this can be seen in Figure 3 at 17:00 UT, particularly in panels d and e.

In total, 645 events (positive classifications) are examined over the 50 dates. Although this technique only returns positive classifications, it is technically a binary classifier given that all spectra without events are predicted "negatives" (i.e contain no Type III emission or bursts that do not meet the threshold). In this way, the true negative can be estimated as the number of spectra with no positive classification. However, the total number of spectra measured in a day relative to the number of spectra with positive classifications is high. Across the 50 sampled dates, on average, 2.8 % of the 3 minute spectra of each day contain an event, although this value is higher during certain periods, such as Type III storms. Such an estimate of the number of true negatives would bias the values and give a disproportionate picture of the classification performance. For these reasons, classification performance metrics based on true negative predictions are not included here. Nonetheless, the number of true negatives is calculated and shown with the number of false negatives in the contingency table in Table 1 for completeness. In the contingency

Table 1: *Contingency table resulting from visual examination of 50 randomly chosen dates. Number of true positives (top left), false positives (bottom left) and false negatives (top right) only are used to evaluate the performance due to the much larger number of true negatives; calculated automatically from the number of spectra with no event present (see text). Counts for the subsets of dates based on the presence of Type III storms are not shown.*

	Predicted positive	**Predicted negative**
Actual positive	255	29
Actual negative	390	22346

Table 2: *Event and classification statistics for the random sample of dates chosen for visual evaluation. The middle and bottom rows refer to subsets of dates excluding Type III storms and considering only Type III storms, respectively. TPR and PPV refer to the true positive rate (recall) and positive predictive value (precision) defined in Equations 1 and 2 respectively. TPR and PPV values in the top row are calculated from values shown in Table 1.*

	N_{dates}	N_{events}	**TPR**	**PPV**
All	50	645	0.90	0.40
Without storms	39	288	0.97	0.52
Only storms	11	357	0.82	0.30

table, the top left value shows the number of true positive (TP) predictions, the bottom left value shows false positive (FP) predictions, the top right shows false negatives (FN) and the bottom right shows true negatives.

With the TP, FN and FP numbers, two important binary classification statistics can be determined. These are the true positive rate (TPR), or recall, and the positive predictive value (PPV), or precision of the Type III selection. The recall is given by the following equation

$$TPR = \frac{TP + FN}{P} \tag{1}$$

where P is the number of "real" Type III bursts observed by Wind, confirmed by visual examination. The TPR thus defines the ability of the selection to return a given spectrum where the beginning of a Type III is present. The precision is given by

$$PPV = \frac{TP}{TP + FP} \tag{2}$$

and thus defines the fraction of classifications that are true Type III bursts. Given the aforementioned effect of many bursts or Type III storms, days where these are observed are noted and the statistics are shown in Table 2 for all dates, those excluding storms and just those including storms.

The selection is shown to return the majority of the observed Type III bursts, as described by the recall value of 0.90 when considering all events in the sampled dates. The aforementioned effect of the Type III storms is seen given the higher recall value of 0.97 for days where they are discounted. The precision is relatively low, 0.40 for all observations, and 0.52 when excluding Type III storms. This suggests that, in both cases, not all classifications define true Type III bursts (as 0.60 of all classifications in the random sample are not visually identified as true Type III bursts) and the classifications of this technique are more precise under periods without Type III storms. This may lead to less precise classification during times of higher solar activity. Although the precision is low, the high recall suggests that this event selection is an effective first step towards selecting Type III events over the Wind dataset, particularly given that the goal here is to select the initial times of Type III bursts. While confirmation may be needed of individual bursts returned by this selection, the high recall shows the reliability of this list as the basis of statistical studies.

In the future, days with the presence of Type III storms and thus a poorer classification could be flagged by calculating the daily average of integrated power between 500-1045 kHz, a frequency range that characterises emission from Type III storms, using filtered data as in panels d and e of the examples in Figures 2 and 3. Comparing this value with a longer term average where fewer Type III storms are observed, over a month or a year and perhaps outside of solar maximum for example, would indicate the presence of a storm if the ratio of daily to longer term averages is higher. The event list shown here could also be compared with other Type III catalogues to discount false positives. However, the discrepancy between frequencies sampled by different instruments and the observing constraints that arise must be considered. This is described in the following section.

3.3 Comparison with the Zhang et al. (2018) NDA catalogue of Type III bursts

The NDA is a decametric radiotelescope hosted at the Nançay radio observatory. Its 144 helical antenna correspond to a $\sim$7000 m^2 effective area at 25 MHz. The NDA quasi-daily observes the Sun in the 10-80 MHz band since Jan. 1978 with various receivers (Boischot et al., 1980; Lamy et al., 2017). Using NDA solar observations acquired with the Routine Receiver (at 0.5 sec $\times$ 175 kHz resolution) between 2012 and 2017, Zhang et al. (2018) inferred a list of type-III bursts. Using the image created by the frequency-time spectrograms, their method uses an algorithm with various transforms and image processing techniques such as contour tracking to identify times and spectral structure of the burst. The returned information includes the observed time of the Type III (defined as the time of the highest frequency emission of the automatically selected burst) and frequency bounds of the emission; the catalogue will be referred to here as the Z18 catalogue. Each of the events in the Z18 catalogue are compared with the selection here to determine the time of the closest event observed by Wind/WAVES; it must be noted that the NDA and Wind/Waves/RAD1 observe a different spectral range, with a different sensitivity and time-frequency resolution, preventing from detailed comparisons. Once matched with a Wind event, the Z18 events are then discounted if the closest Wind burst is closer to another Z18 event. This assumes that each Type III burst observed by NDA above 10 MHz will correspond to a single Type III burst in the Wind observations, thus evaluating the temporal precision of the selection relative to the NDA events. The time difference $\Delta t = t_W - t_{Z18}$ between matching events is then calculated, where t_W is the time returned by the technique of this paper and t_{Z18} is the start time of the event published in Z18. The average drift rate is given in the Z18 catalogue; while this could be used to estimate an expected observation time of the emission at 1040 kHz (the highest observing frequency of the Wind/WAVES/RAD1 receiver), the drift rate is expected to

Table 3: *Percentile values for Δt, in minutes, computed over 967 Z18 bursts with unique matches in the Wind event list.*

Percentiles	Lower percentile (minutes)	Higher percentile (minutes)
0th (min) - 100th (max)	-5205.	8651
5th - 95th	-48	68
10th - 90th	-24	25
20th - 80th	-8	9

change as the frequency decreases (Reid & Ratcliffe, 2014). As the drift rate is calculated as an average between the frequency bounds of a given Type III, which vary between the observing range of 10-80 MHz, we use the observation time from the Z18 catalogue to calculate the difference.

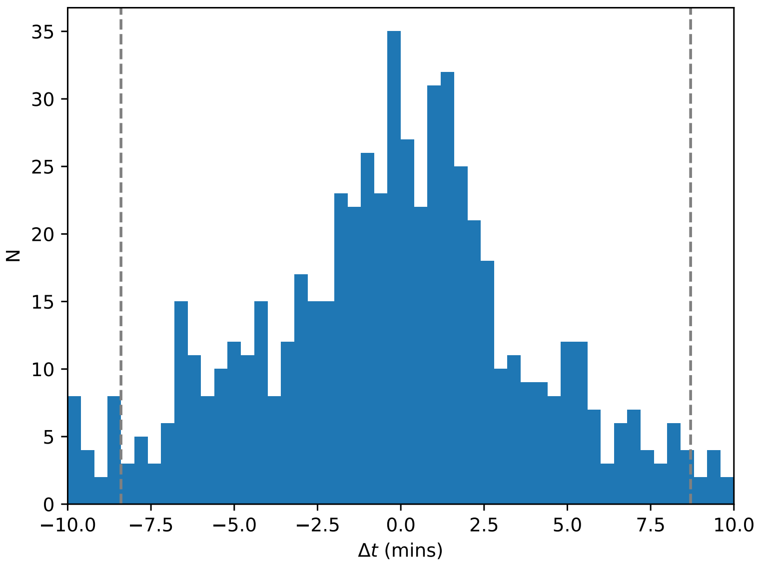

Figure 5: *Distribution of Δt, the time difference in minutes between Z18 and Wind Type III bursts where a unique match is found. The distribution does not show the entire range of data; between 60% and 80% of the matches are shown for clarity. The extremes of the distribution are given in Table 3.*

The distribution of Δt in minutes, for Z18 bursts with a unique corresponding event in this event list, are shown in Figure 5. The plot limits are chosen based on the 20^{th} and 80^{th} percentiles of the data to for greater clarity. ~ 70 % of Z18 bursts (967 of 1389) have unique matches with the event list here, while ~ 63 % of those have a corresponding match within $\pm$ 10 minutes. These matches account for approximately 0.04 % of Wind events that occur within the corresponding NDA Type III observation window. The minimum and maximum Δt values (0^{th} and 100^{th} percentiles), as well as the 5^{th}-95^{th} and 10^{th}-90^{th} percentiles of Δt are shown in Table 3.

Given that NDA observes above 10 MHz, there is an expected temporal discrepancy between the Z18 and Wind events. The drifting morphology of Type III bursts, with an increasing drift rate at lower frequencies and particularly below 1 MHz, suggests that corresponding events observed by Wind/WAVES/RAD1 would occur later than those by NDA. However, the nature of the selection here, where a fixed threshold limit is used, is

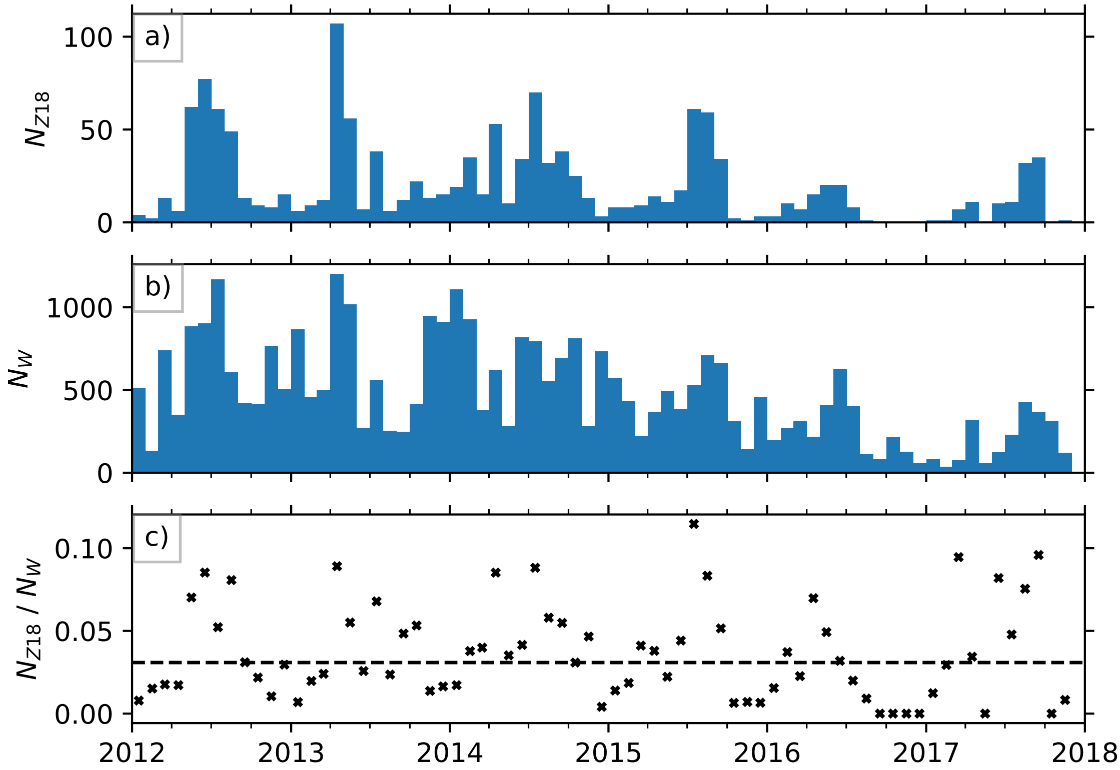

Figure 6: *Event occurrence of (a) all Type III bursts from Z18, (b) all bursts from Wind and (c) the ratio $\frac{N_{Z18}}{N_W}$ from 2012 to 2017, shown with monthly bins. The horizontal dashed line in the bottom panel shows the median ratio for each month (0.031).*

such that the time indicating the Type III burst could be earlier than that of the most intense part of the burst. This would thus be expected comparing to the Z18 events, where the brightest portion of the dynamic spectrum defines the Type III burst. As shown in Figure 5, these effects likely balance the distribution.

Another effect to consider is that of the scattering of lower frequency emission from density inhomogenities within the solar plasma environment out to 1 AU (Krupar et al., 2020; Musset et al., 2021). As well as directivity effects in general (Bonnin et al., 2008), density fluctuations and the consequential scattering can beam Type III emission from sources at a greater range of solar longitudes (outside of the Sun-Earth line) at Wind observing frequencies than those of NDA (Steinberg et al., 1971; Kontar et al., 2017, 2019), given that lower frequency emission originates from electrons further from the Sun. Although the resulting emission would be weaker, this means that Wind would be expected to observe a greater number of Type IIIs that would not correspond directly to events in the Z18 catalogue. As this is indistinguishable in a statistical sense from the effect of the false positive classifications described in Section 3.2, the effect due to this observing discrepancy would require further study to constrain.

The number of events returned by the Z18 catalogue and the Wind selection here are shown in Figure 6, for the entire period covered by the Z18 catalogue. Events are binned in month-long bins, with the Z18 events (N_{Z18}), the Wind events (N_W) and their ratio ($\frac{N_{Z18}}{N_W}$)

shown in panels a-c. Wind events in the middle panel are not limited to unique matches with Z18; all events are shown. There is broad correlation between the Z18 catalogue and the Wind list. However, the Wind list consistently returns at least $\sim$10$\times$ the number of events as seen in the bottom panel, with a median ratio value of 0.031 (represented by the dashed line). This is likely due to a combination of the aforementioned effects; Wind will observe emission from a wider range of solar longitudes due to greater scattering at low frequencies, as well as a number of false positives based on the precision values in Table 2. (Raja et al., 2022) report on the median intensity spectrum of a statistical sample of Type III bursts, showing that the median intensity across the integrated power range considered by Wind here (500-1040 kHz) is greater than that for the NDA observation frequencies (10-80 MHz). However, the sensitivity of NDA ($\sim$7000 m^2) orders of magnitude larger than that of the Z antenna (77 m^2, estimated from the squared effective antenna length $L_{eff} \sim 8.78\ m$ (Bonnin et al., 2023)). The discrepancy between returned events is thus unclear and is not explored here in detail, although the broad correlation strengthens the use of this event list for statistical studies.

4 Discussion and future work

By using the inverse of a binary mask developed to extract highly spin-variable AKR observations from Wind with a SNR selection, the method presented here introduces an automatic selection of Type III solar radio bursts from Wind/WAVES/RAD1 observations. The selection automatically removes observations where AKR is the dominant emission, so Type III bursts observed simultaneously with AKR are discounted; usually occurring when Wind is on the flanks of the magnetosphere or the nightside.

The selection is applied to 27 years of Wind data between 1995-2021, integrating the power between 500-1045 kHz and using a threshold of 1×10^9 W sr^{-1}. This period covers two solar cycles; comparison of the event occurrence with the sunspot number in Figure 4 shows good correlation, with a Pearson coefficient of 0.94 obtained when calculated with yearly averaged sunspot number. Quantifying the selection performance by visual examination of events during a random sample of 50 dates, chosen across the solar cycle and accounting for different spacecraft local times, returns a recall (TPR) of 0.9 as seen in Table 2. The recall increases to 0.97 when discounting dates with Type III storms present, showing that the selection returns the majority of Type III bursts but is less effective when the daily integrated power above 500 kHz is increased, requiring a different threshold to properly track the variability. While most Type IIIs are returned some classifications are spurious, as shown by the precision (PPV) of 0.4 for all dates and 0.52 when discounting days with Type III storms. While the observing frequencies of Wind/WAVES/RAD1 and the ground-based NDA are much different, comparing the event lists in Section 3.3 showed that approximately 70% of NDA bursts in the Z18 catalogue have a unique match in the Wind list, with 63% of those occurring within $\pm$10 minutes. The Wind catalogue returns $\sim$33$\times$ the number of events for the period shown by Zhang et al. (2018) based on the median value of the occurrence ratio in Figure 6, and shows broad correlation with their event occurrence.

The precision of the selection could be improved by choosing a value based on the relative power increase above the daily average for a threshold, instead of a fixed value. As well as this, a statistical parameterisation of the σ_z threshold could be implemented to more carefully select Type IIIs, and refine the spectral characterisation to explore the lower frequency, fainter parts of the bursts. The SNR threshold could also be varied easily if more or less sensitivity is required. The concept of this technique could also be possible to apply to the Wind/WAVES/RAD2 receiver, which operates between 1-13 MHz where AKR is not present. As it is also possible for the RAD1 receiver to observe Type II solar emission, the breadth of Wind observations could be utilised to explore these events if possible to be selected by this technique, depending on the typical σ_z values of Type II bursts.

Similarly, while the use of integrated power can simply give the times of the first observation of a Type III burst, more sophisticated tools could be used to return the full temporal and spectral information of the Type III burst. The SPACE labelling tool (Louis et al., 2022b; Louis et al., 2022a), which allows the user to select polygons on frequency-time dynamic spectrograms and return TFCat-compliant (Cecconi et al., 2023) data to store the bounding information of the Type III. However, using the SPACE labelling tool to categorise Type III emission observed from L1 by Wind, with 20 years of data at present, would be manually intensive. Techniques such as k-means clustering could be used instead to automatically select the bright regions of Type III bursts from the outputs seen in figures 2d and 3d. While the integrated power threshold as described above could be simply implemented in the first instance, such additional steps could improve the precision of the selection in the future. Not only can previous observations be explored statistically, such as from the quasi-constant position of Wind and preliminary work over multiple solar cycles in Figure 4, but the potential for near real time application of this technique could prove useful in a space weather monitoring context. Such an algorithm could be used for alert systems, to prompt other, higher resolution instruments to begin recording, or to pre-process onboard acquisitions for a future mission, for example.

Acknowledgements

The authors acknowledge Centre National d'Etudes Spatiales (CNES), Centre National de la Recherche Scientifique (CNRS), and Observatoire de Paris for support to the Wind/Waves team and the Centre de Données de la Physique des Plasmas (CDPP) for the provision of the Wind/Waves RAD1 L2 data. The authors also acknowledge support from CNRS/INSU national programs of planetology (PNP) and heliophysics (PNST). The authors acknowledge support from Paris Astronomical Data Centre (PADC) for the preparation and distribution of the data collection. CMJ's, CKL's and ARF's work is supported by the Science Foundation Ireland Grant 18/ FRL/6199. ARF's work was also supported by Irish Research Council Government of Ireland Postdoctoral Fellowship GOIPD/2022/782. DKW was supported by NERC IRF NE/S015167/1. The authors thank the Solar Influences Data Analysis Center (SIDC) at the World Data Center SILSO, Royal Observatory of Belgium, Brussels, for providing corrected, averaged sunspot numbers.

References

Boischot A., et al., 1980, A new high-grain, broadband, steerable array to study Jovian decametric emission, *Icarus*, *43, 399*

Bonnin X., Hoang S., Maksimovic M., 2008, The directivity of solar type III bursts at hectometer and kilometer wavelengths : Wind-Ulysses observations, *Astronomy & Astrophysics*, *427, 419*

Bonnin X., Hoang S., Cecconi B., Issautier K., 2023, Description of Wind/Waves/RAD1 L3 Direction-Finding data production [dataset] Version 1.0, doi:10.25935/4ztj-0s14

Bougeret J. L., et al., 1995, WAVES: The radio and plasma wave investigation on the wind spacecraft, *Space Science Reviews*, *71, 231*

Cecconi B., Zarka P., 2005, Direction finding and antenna calibration through analytical inversion of radio measurements performed using a system of two or three electric dipole antennas on a three-axis stabilized spacecraft, *Radio Science*, *40, RS3003*

Cecconi B., et al., 2008, STEREO/Waves Goniopolarimetry, *Space Science Reviews*, *136, 549*

Cecconi B., Louis C. K., Bonnin X., Loh A., Taylor M. B., 2023, Time-frequency catalogue: JSON implementation and python library, *Frontiers in Astronomy and Space Sciences*, *9*

Cliver E. W., Ling A. G., 2009, Low-frequency type III bursts and solar energetic particle events, *Astrophysical Journal*, *690, 598*

Dulk G. A., 2000, Type III Solar Radio Bursts at Long Wavelengths, *Geophysical Monograph Series*, *119, 115*

Eastwood J. P., Wheatland M. S., Hudson H. S., Krucker S., Bale S. D., Maksimovic M., Goetz K., Bougeret J. L., 2010, On the brightness and waiting-time distributions of a type III radio storm observed by stereo/waves, *Astrophysical Journal Letters*, *708, 95*

Fogg A. R., Jackman C., Waters J., Bonnin X., Lamy L., Cecconi B., Issautier K., Louis C. K., 2021, Bursts of auroral kilometric radiation individually selected from wind/waves data (version 1.0) [data set], *MASER/PADC*,

Fogg A. R., Jackman C. M., Waters J. E., Bonnin X., Lamy L., Cecconi B., Issautier K., Louis C. K., 2022, Wind/WAVES Observations of Auroral Kilometric Radiation: Automated Burst Detection and Terrestrial Solar Wind–Magnetosphere Coupling Effects, *Journal of Geophysical Research: Space Physics*, *127, e2021JA030209*

Hanasz J., Schreiber R., Aksenov V. I., 1980, Polarized solar type III bursts between 2.3 and 4.9 MHz, *Astronomy & Astrophysics*, *91, 311*

Hoang S., Steinberg J. L., Reiner M. J., Stone R. G., 1995, Interplanetary type iii radio bursts observed simultaneously by ice, *Space Science Reviews*, *72, 267*

Kontar E. P., et al., 2017, Imaging spectroscopy of solar radio burst fine structures, *Nature Communications*, *8, 1515*

Kontar E. P., et al., 2019, Anisotropic radio-wave scattering and the interpretation of solar radio emission observations, *The Astrophysical Journal*, *884, 122*

Krupar V., et al., 2018, Interplanetary Type III Bursts and Electron Density Fluctuations in the Solar Wind, *The Astrophysical Journal*, *857, 82*

Krupar V., et al., 2020, Density fluctuations in the solar wind based on type iii radio bursts observed by parker solar probe, *The Astrophysical Journal Supplement Series*, *246, 57*

Lamy L., Zarka P., Cecconi B., Klein L., Masson S., Denis L., Coffre A., Viou C., 2017, 1977-2017: 40 years of decametric observations of Jupiter and the Sun with the Nancay Decameter Array, *in Planetary Radio Emissions VIII, eds Fischer, G. and Mann, G. and Panchenko, M. and Zarka, P.*, pp 455–466, doi:10.1553/PRE8s455

Lobzin V. V., Cairns I. H., Zaslavsky A., 2014, Automatic recognition of type III solar radio bursts in STEREO/WAVES data for onboard real-time and archived data processing, *Journal of Geophysical Research: Space Physics*, *119, 742*

Louis C. K., et al., 2022a, SPACE Labelling Tool Version 2.0.0 (v2.0.0) [Code], doi:10.5281/zenodo.6886528, `https://doi.org/10.5281/zenodo.6886528`

Louis C., et al., 2022b, The "SPectrogram Analysis and Cataloguing Environment" (SPACE) labelling tool, *Frontiers in Astronomy and Space Sciences*, *9*

Musset S., et al., 2021, Simulations of radio-wave anisotropic scattering to interpret type III radio burst data from Solar Orbiter, Parker Solar Probe, STEREO, and Wind, *Astronomy and Astrophysics*, *656*

Pelton J. N., Allahdadi F., 2015, Handbook of Cosmic Hazards and Planetary Defense. Springer International Publishing, Cham, doi:10.1007/978-3-319-03952-7, `http://link.springer.com/10.1007/978-3-319-03952-7`

Potter D. W., Lin R. P., Anderson K. A., 1980, Impulsive 2-10 kev solar electron events not associated with flares, *The Astrophysical Journal*, *236, L97*

Raja K. S., et al., 2022, Spectral analysis of solar radio type iii bursts from 20 khz to 410 mhz, *The Astrophysical Journal*, *924, 58*

Reid H. A. S., Ratcliffe H., 2014, A review of solar type iii radio bursts, *Research in Astronomy and Astrophysics*, *14, 773*

Reiner M. J., Fainberg J., Kaiser M. L., Bougeret J. L., 2007, Circular polarization observed in interplanetary type III radio storms, *Solar Physics*, *241, 351*

Saint-Hilaire P., Vilmer N., Kerdraon A., 2013, A decade of solar type III radio bursts observed by the Nançay Radioheliograph 1998-2008, *Astrophysical Journal*, *762*

Scully J., Flynn R., Gallagher P. T., Carley E. P., Daly M., 2023, Improved type iii solar radio burst detection using congruent deep learning models, *Astronomy & Astrophysics*, *674, A218*

Singh D., Sasikumar Raja K., Subramanian P., Ramesh R., Monstein C., 2019, Automated Detection of Solar Radio Bursts Using a Statistical Method, *Solar Physics*, *294*

Steinberg J. L., Aubier-Giraud M., Leblanc Y., Boischot A., 1971, Coronal Scattering, Absorption and Refraction of Solar Radiobursts, *Astronomy & Astrophysics*, *10, 362*

Stewart R. T., 1974, Harmonic ratios of inverted-U type III bursts, *Solar Physics*, *39, 451*

Waters J., Cecconi B., Jackman C. M., Whiter D. K., Bonnin X., Lamy L., 2021a, Wind/waves flux density collection calibrated for auroral kilometric radiation (version 1.0) [data set], *PADC*,

Waters J. E., Jackman C. M., Lamy L., Cecconi B., Whiter D. K., Bonnin X., Issautier K., Fogg A. R., 2021b, Empirical Selection of Auroral Kilometric Radiation During a Multipoint Remote Observation With Wind and Cassini, *Journal of Geophysical Research: Space Physics*, *126, e2021JA029425*

Waters J. E., et al., 2022, A Perspective on Substorm Dynamics Using 10 Years of Auroral Kilometric Radiation Observations From Wind, *Journal of Geophysical Research: Space Physics*, *127*

Wild J., 1950, Observations of the Spectrum of High-Intensity Solar Radiation at Metre Wavelengths. III. Isolated Bursts, *Australian Journal of Chemistry*, *3, 541*

Wild J. P., Smerd S. F., Weiss A. A., 1963, Solar Bursts, *Annual Review of Astronomy and Astrophysics*, *1, 291*

Zarka P., Cecconi B., Kurth W. S., 2004, Jupiter's low-frequency radio spectrum from Cassini/Radio and Plasma Wave Science (RPWS) absolute flux density measurements, *Journal of Geophysical Research: Space Physics*, *109, A09S15*

Zhang P. J., Wang C. B., Ye L., 2018, A type III radio burst automatic analysis system and statistic results for a half solar cycle with Nançay Decameter Array data, *Astronomy and Astrophysics*, *618, 1*