

Document downloaded from the institutional repository of the University of Alcalá: <http://ebuah.uah.es/dspace/>

This is a preprint version of the following published document:

Larrodera, C. & Cid, C. 2020, "The distribution function of the average iron charge state at 1 AU: from a bimodal wind to ICME identification", Solar Physics, vol. 295, 156 (2020).

Available at <https://doi.org/10.1007/s11207-020-01727-8>

© 2020 Springer Nature

*(Article begins on next page)*



This work is licensed under a

Creative Commons Attribution-NonCommercial-NoDerivatives  
4.0 International License.



material from the solar atmosphere is thrown into the interplanetary space (Schwenn, 2006b; Viall and Borovsky, 2020). Each type of bulk wind has a different source. The fast solar wind is considered to come from the open magnetic field that emerges from the Sun through the coronal holes (Banaszkiewicz et al., 1997; Schwenn, 2006a). The source of the slow solar wind is not clear, and many suggestions have been made, e.g. plasma released by reconnection between open and closed magnetic field lines (Lionello et al., 2005) or flow emerging from small equatorial coronal holes (Bale et al., 2019). Nevertheless, Neugebauer et al. (2002), from a study of the sources of the solar wind during the maximum of Solar Cycle 23 (1998-2001), conclude that near the solar maximum the characteristics of the fast solar wind are very different from the fast solar wind during the minimum when the polar coronal holes are dominant because of the contribution of the active regions to the fast stream. They suggest a hierarchy based on open field regions with large polar coronal holes as the source of the highest speeds, and active regions like other contributors to the fast wind showing a correlation between the fast wind and the solar cycle.

Through the years different distribution functions have been used to model different solar wind parameters, e.g. a lognormal function (Burlaga and King, 1979) or kappa-like (Vörös et al., 2015) for the magnetic field, or a gamma-like function (Li, Zhanng, and Feng, 2016) for solar wind speed, among others. Recently Larrodera and Cid (2020) proposed the bi-Gaussian function, as a characterization of the solar wind at 1 AU. This model has been applied to the distribution function, not only of the proton speed, but also of the proton temperature, magnetic field and density, showing a bimodal distribution of all the four solar wind parameters.

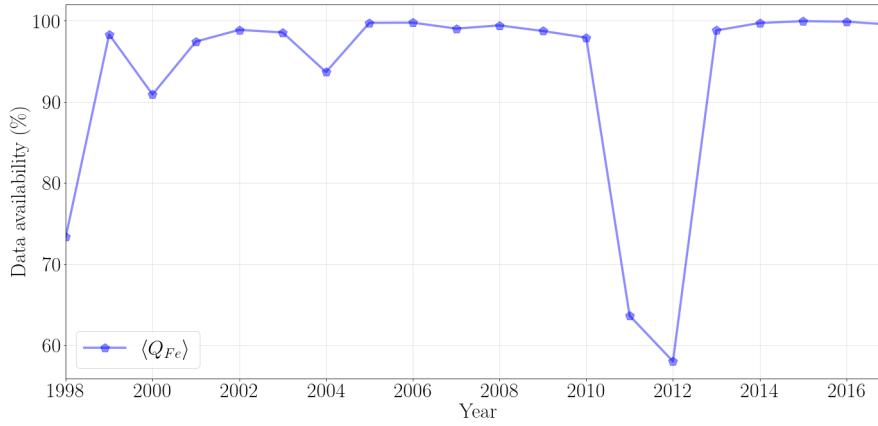
While the solar wind speed or the proton density can dynamically evolve between the Sun and the Earth as a result of the interactions of the different streams coming out from the Sun, no major changes are expected in the composition of the solar wind, which is frozen since its departure from the Sun. Thus, the solar wind composition can be considered a useful physical magnitude to characterize it (Cranmer, Gibson, and Riley, 2017; Schwenn, 2006a). Indeed, the solar wind composition is used as a signature of fast streams coming from coronal holes (e.g. Heidrich-Meisner et al. (2016)) and interplanetary coronal mass ejections (ICMEs). In particular, the average iron charge state ( $\langle Q_{Fe} \rangle$ ), has been used to identify ICMEs (Lepri et al., 2001; Lepri, 2004) or to establish their boundaries (Cid et al., 2016).

In this article we study the distribution function of  $\langle Q_{Fe} \rangle$  measured by the Solar Wind Ion Composition Spectrometer (SWICS) instrument (Gloeckler et al., 1998) on board the Advance Composition Explorer (ACE). Our first aim is to check if the bi-Gaussian function is able to separate each contribution from the slow and fast winds. We also study the relevance of ICMEs in the solar wind sample and how the bimodal approach is related with the solar cycle. Section 2 describes the data set used in the study. Section 3 applies the bi-Gaussian approach to the distribution function of  $\langle Q_{Fe} \rangle$  at 1 AU, first to the whole data sample and then to a reduced data sample, after removing ICMEs and outliers from the bulk solar wind. Then, in Section 4 we analyze the outliers of the sample, comparing to previously identified ICMEs. Finally, Section 5 details our conclusions.

## 2. Data

This study uses data from SWICS on board ACE. Specifically, we use level 2 average iron charge state data with two-hour resolution from the ACE Science Data Center (<http://www.srl.caltech.edu/ACE/ASC/level2/index.html>).

Figure 1 shows the data availability during the period under study (1998-2017). The drop between 2010-2013 was due to a radiation and age-induced hardware anomaly which altered the instrument operational state, as informed by the operational SWICS team. Nevertheless the lowest availability value is around 60 % which is high enough for the analysis.



**Figure 1.** Data availability of  $\langle Q_{Fe} \rangle$  measured by SWICS/ACE.

Indeed, there are two different SWICS data sets: SWICS 1.1, prior to the 23<sup>th</sup> August 2011 anomaly, and SWICS 2.0, for the time period after the anomaly. According to the Data Release Notes from the instrument team, SWICS 1.1 data provide a reliable ground truth for validation of methods to obtain SWICS 2.0 data. Therefore, we have checked the results obtained from the whole SWICS data set against those obtained only using SWICS 1.0 data. Although the data in this sample is reduced in this case, we consider that the results obtained from the whole SWICS data set are robust (and therefore described in this article) if they coincide with those from SWICS 1.0 data set.

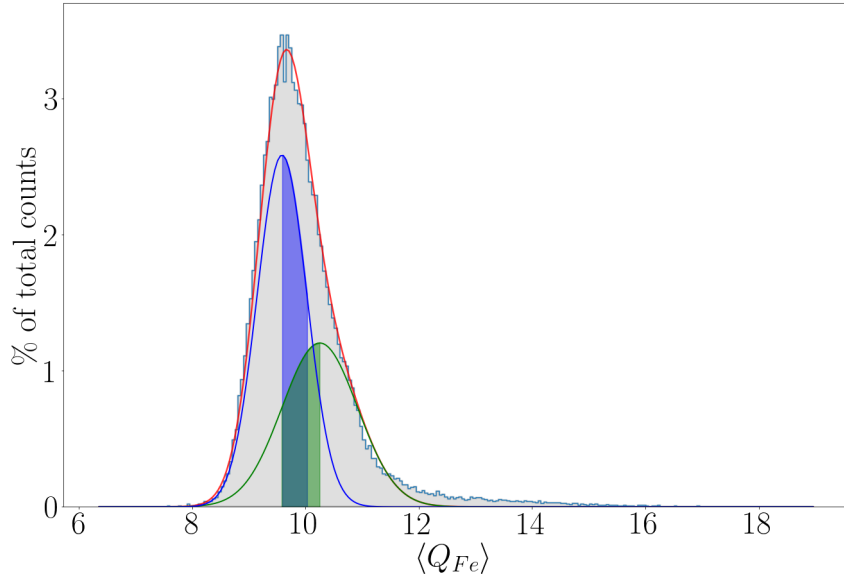
## 3. A Bi-Gaussian approach for $\langle Q_{Fe} \rangle$

We propose a bi-Gaussian function as the probability distribution function (PDF) for  $\langle Q_{Fe} \rangle$  at 1 AU. This function is defined as the addition of two Gaussian distribution functions where each one represents the contributions of one type of solar wind. A bi-Gaussian function has been previously used to explain the distribution of other solar wind magnitudes (Larrodera and Cid, 2020).

The analytical expression of the bi-Gaussian function in this case is:

$$bG(\langle Q_{Fe} \rangle) = h_1 \cdot \exp\left(\frac{-\langle Q_{Fe} \rangle - p_1)^2}{2w_1^2}\right) + h_2 \cdot \exp\left(\frac{-\langle Q_{Fe} \rangle - p_2)^2}{2w_2^2}\right) \quad (1)$$

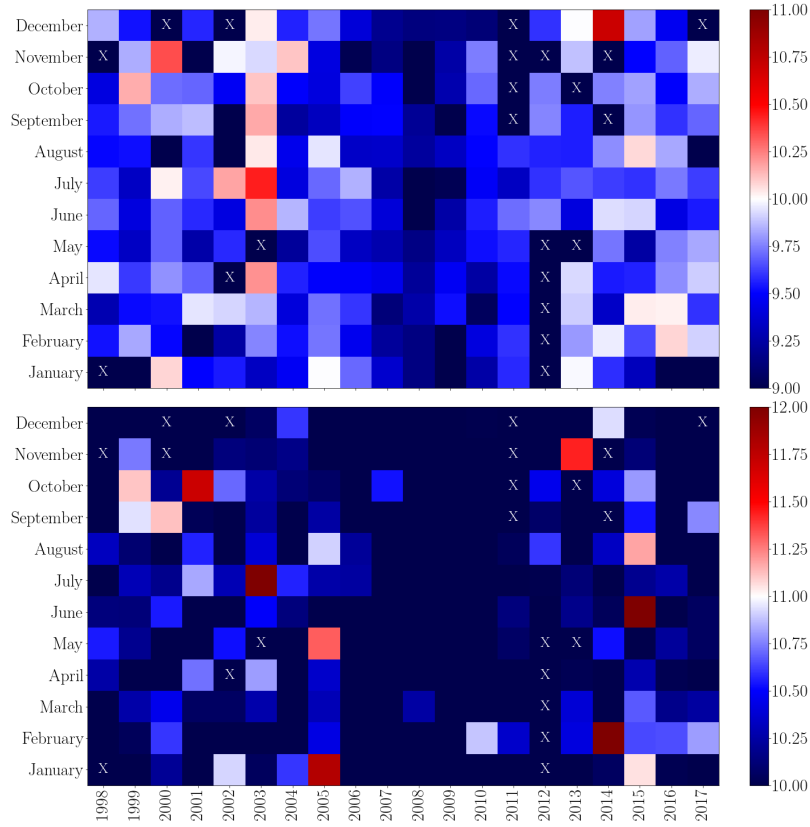
where  $h_1$ ,  $p_1$ ,  $w_1$ ,  $h_2$ ,  $p_2$  and  $w_2$  are the parameters obtained from the fitting to the data set. The subscripts of these parameters correspond to the first (1) or second (2) Gaussian distribution function, respectively, representing the two types of wind.  $h$  represents the height of the peak of each single Gaussian,  $p$  the position of the peak, and  $w$  the root-mean-square (RMS) width of each single Gaussian.



**Figure 2.** Empirical distribution function of  $\langle Q_{Fe} \rangle$  for the whole data set under study from SWICS/ACE (grey) and the fitting to a bi-Gaussian function (red). Green and blue lines correspond to the single Gaussian curves. See text for explanation on blue and green shadowed areas.

Figure 2 shows the normalized distribution of the whole data set of  $\langle Q_{Fe} \rangle$  and the fitting to a bi-Gaussian function. Blue and green lines represent the individual Gaussian functions for each type of wind. Most of  $\langle Q_{Fe} \rangle$  is spread around  $p_1 = 9.6$  but there is also another significant contribution around  $p_2 = 10.3$ . The Pearson- $\chi^2$  value from the fitting is 4.4 with a data availability of 93.3%, showing that a bimodal distribution can be considered. Nevertheless, assuming  $w_1 = 0.4$  and  $w_2 = 0.7$  as the uncertainties of  $p_1$  and  $p_2$ , respectively, we obtain that the position of both peaks overlaps. This is shown by shadowed regions: the blue region covers the interval  $(p_1, p_1 + w_1)$  and the green region the  $(p_2 - w_2, p_2)$  one. The small difference between the position of the center of the peak of each single Gaussian questions the bi-Gaussian approach for  $\langle Q_{Fe} \rangle$ . Indeed, yearly samples might have missed a detailed behaviour which is expected to arise in

shorter samples. Thus we study the monthly evolution of  $\langle Q_{Fe} \rangle$  distribution function with the aim of solving this concern.

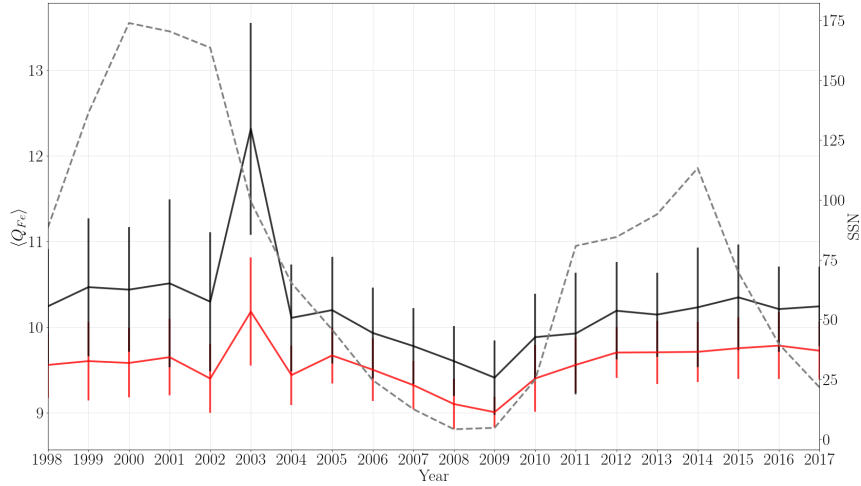


**Figure 3.** Positions  $p_1$  and  $p_2$  from the bi-Gaussian fitting of the monthly  $\langle Q_{Fe} \rangle$  distribution functions. Top (bottom) represents the first (second) peak. The white 'Xs' marks the months where the fitting was not possible.

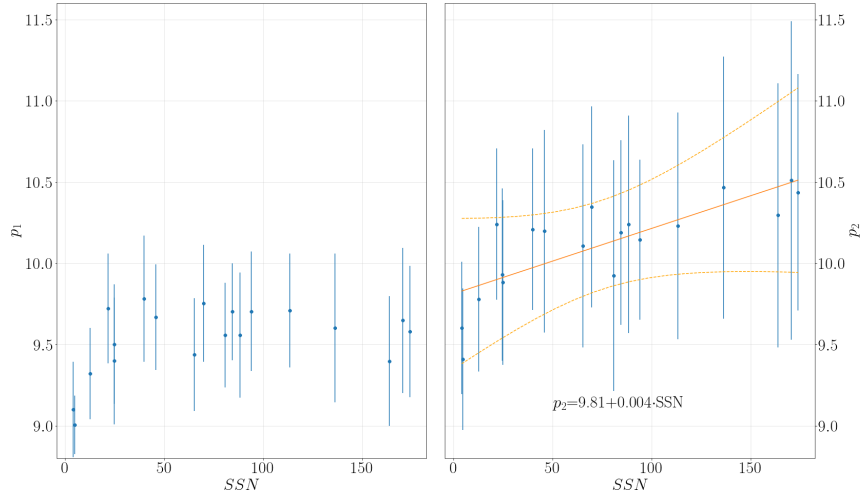
Figure 3 represents the position of the peak of the first (top) and second (bottom) Gaussian functions for each month along the whole data set. Note that the range of the color scale in both figures is slightly different. We observe three vertical bands which match the periods of the solar cycle: (1) from the ascending to the declining phase of Solar Cycle 23 (1998-2005), (2) during the minimum (2006-2012), and (3) from the ascending to the declining phase of Solar Cycle 24 (2013-2017).

Considering this result, we study the relationship between the bimodal  $\langle Q_{Fe} \rangle$  distribution function and the solar cycle, using the sunspot number (SSN) as a proxy for solar activity. Figure 4 compares the position of the peaks from the bi-Gaussian fitting and the SSN over time. Year 2003 diverges from the trend of the rest of the period analyzed because of a very high activity with a major contribution from the fast wind. Indeed, Larrodera and Cid (2020) described it as an unusual year dominated by fast

solar wind and a large number of CMEs. Therefore, we have avoided including year 2003 to search for any relationship with the solar cycle. Figure 5 shows the scatter plots of positions  $p_1$  and  $p_2$  versus SSN (excluding year 2003). We cannot perceive any clear correlation between  $p_1$  and the SSN, but  $p_2$  clearly increases as SSN increases. Indeed, a Pearson correlation coefficient  $r(p_2, SSN) = 0.77$  (excluding year 2003) indicates a strong linear correlation between  $p_2$  and the SSN. On the contrary,  $r(p_1, SSN) = 0.40$ , confirms the weak correlation observed between  $p_1$  and the SSN.



**Figure 4.** Position of the peaks from the bi-Gaussian fitting for each year. Red (black) line represents the first (second) Gaussian. The RMS width of the corresponding Gaussian ( $w$ ) has been considered as the uncertainty in the  $p$  parameter. The grey dashed line corresponds to the yearly SSN.



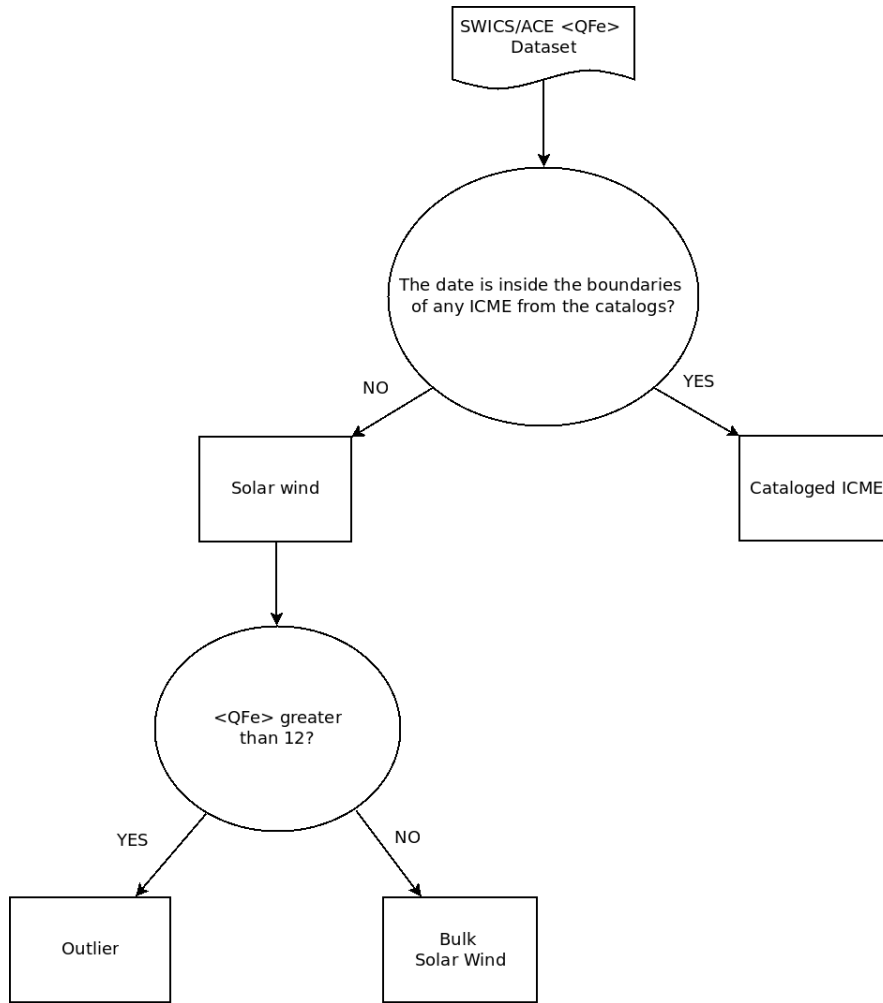
**Figure 5.** Scatter plots of  $p_1$  (left) and  $p_2$  (right) vs SSN. In the right panel, the orange solid line corresponds to the linear regression and the orange dashed lines are the 99% confidence intervals.

### 3.1. Removing Outliers. Analyzing the Bulk Solar Wind

The previous section analyzes the distribution function of the solar wind as a whole data set with all data contributing to the bulk solar wind. However, we appreciate in Figure 3 some months with extremely high values of  $\langle Q_{Fe} \rangle$ , which may be associated with transient events. Indeed, Lepri et al. (2001) and Lepri (2004) considered large  $\langle Q_{Fe} \rangle$  as a sufficient signature of ICMEs. Moreover, Figure 1 in Lepri (2004) shows that the probability that one would find a certain average charge state in the normal solar wind is zero from  $\langle Q_{Fe} \rangle = 12$  on. Checking the results from the previous bi-Gaussian fitting to the whole data set (Figure 2), we notice that  $\langle Q_{Fe} \rangle = 12$  corresponds to  $p_2 + 3\sigma_2$ , and therefore values of  $\langle Q_{Fe} \rangle > 12$  can be considered as outliers of the bulk solar wind. As our first goal is to study the bulk solar wind, we proceed to remove all these outliers from the whole data sample.

We also remove from the whole data sample the ICMEs previously identified. For this purpose we considered the events listed in the following ICME catalogs: Richardson and Cane (<http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>), Jian et al. (2006), Jian, Russell, and Luhmann (2011) and NASA Wind catalog (<https://wind.nasa.gov/cycle22.php>). Figure 6 shows the workflow of the process.

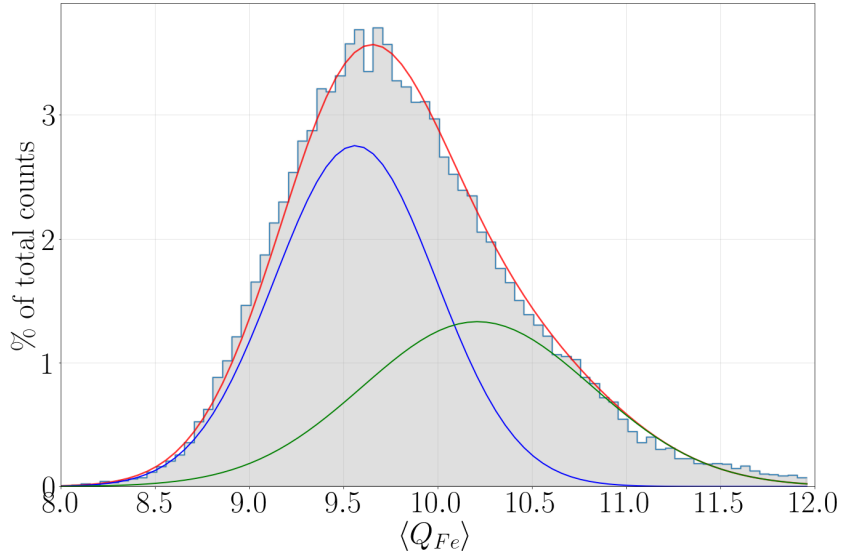




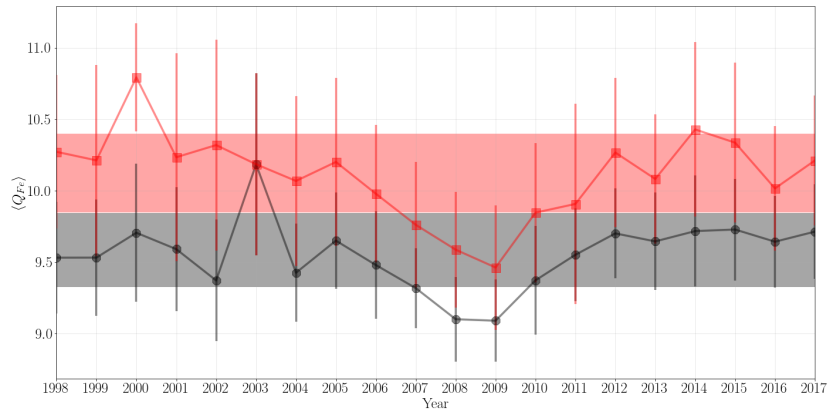
**Figure 6.** Workflow showing the process to separate the bulk solar wind data from the outliers and the ICMEs previously identified.

After this process, the reduced data sample, representing the bulk solar wind, is ready for analysis. Thus, we go ahead fitting the bi-Gaussian function to the PDF of the reduced data set of  $\langle Q_{Fe} \rangle$ , which represent 89 % of the whole.

Figure 7 shows the empirical distribution function of the  $\langle Q_{Fe} \rangle$  for the reduced data sample and the fitting to the bi-Gaussian function. A Pearson- $\chi^2$  value of 0.01 shows a great fit for the bulk solar wind sample. Now, the values of  $\langle Q_{Fe} \rangle$  are spread around  $p_1 = 9.6$  and  $p_2 = 10.2$ , almost around the same values as before, but with smaller uncertainties.



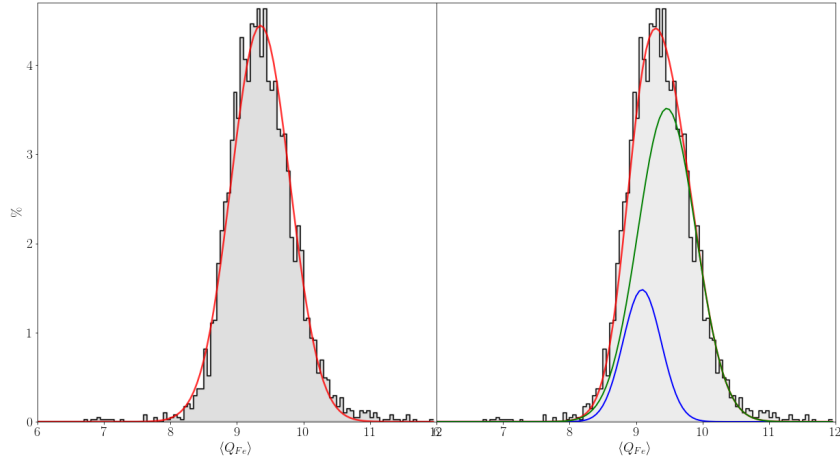
**Figure 7.** Empirical distribution function of the  $\langle Q_{Fe} \rangle$  for the reduced sample (shaded as grey) and fitting to a bi-Gaussian function (red). Green and blue lines correspond to the single Gaussian curves.



**Figure 8.** Position of the peak of the first (black) and second (red) Gaussian distribution function from the yearly fitting. The shadowed areas are the weighted average position  $\pm \sigma_w$ .

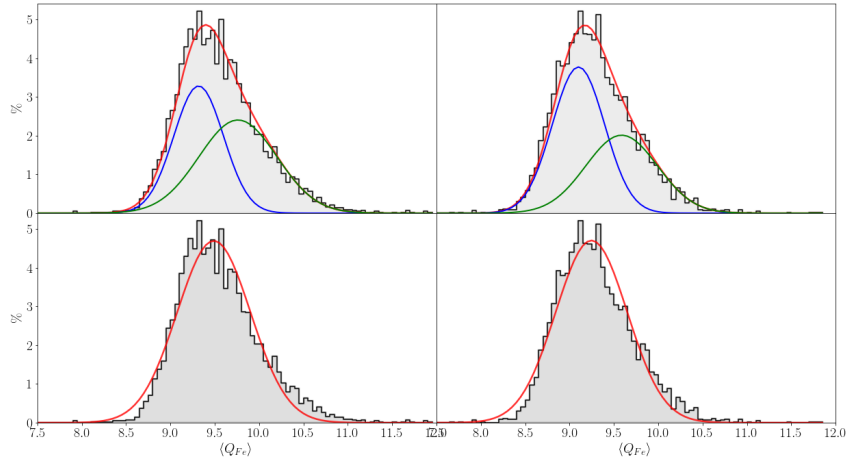
Figure 8 represents the position of the peaks from the bi-Gaussian fitting for each year of the bulk solar wind data set. The red (black) shadowed area corresponds to the weighted average position of the peak  $p_1$  ( $p_2$ )  $\pm \sigma_w$ . Although both colored regions in the plot are separated, reinforcing a bimodal distribution function for  $\langle Q_{Fe} \rangle$  at 1 AU, the position of one of the peaks appears in the region corresponding to the other peak

for some years. This happens, not only for year 2003, which we labeled as anomalous above, but also for years 2007, 2008 and 2009.



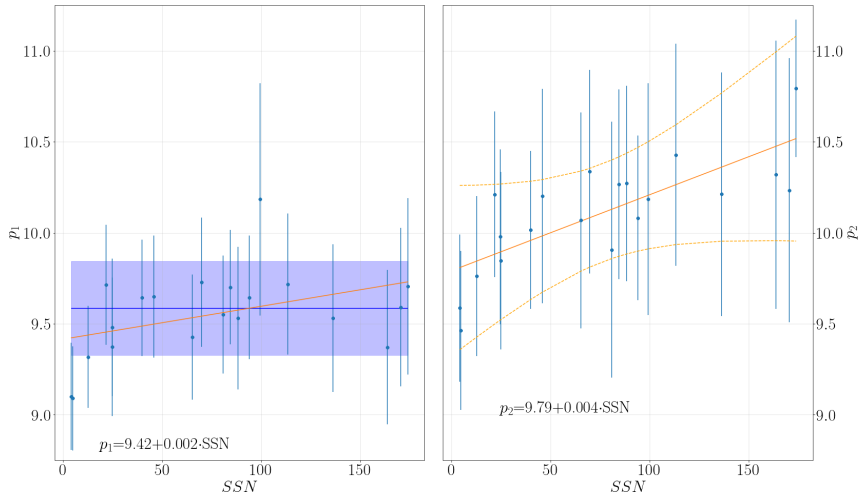
**Figure 9.** Empirical distribution function of the  $\langle Q_{Fe} \rangle$  for the reduced sample of year 2009 with a single Gaussian fitting (left) and bi-Gaussian fitting (right) with the corresponding single curves in blue and green.

Larrodera and Cid (2020) explained that in 2009 the slow solar wind is highly dominant over the fast wind. In this situation, with predominance of slow solar wind, the contribution of the fast solar wind can be discarded and the bi-Gaussian fitting can be replaced by a single Gaussian fitting. This is shown in Figure 9, where we compare the bi-Gaussian and single Gaussian fitting for 2009. The single Gaussian fitting provides a good approximation for the data set, and adding a second one (i.e. using bi-Gaussian fitting) will not provide a better fitting. Indeed, the first Gaussian has a smaller height when compared to the second Gaussian, therefore its contribution can be discarded as this has no physical meaning. Nevertheless, this is not the case for years 2007 and 2008. Figure 10 shows a clear deviation from the single Gaussian fitting of the PDF of years 2007 and 2008 due to the heavy tail. This problem is solved by the bi-Gaussian function. Indeed, the displacement of some years from the corresponding shadowed areas in Figure 8 may be related to the solar cycle.



**Figure 10.** Empirical distribution function of the  $\langle Q_{Fe} \rangle$  for the reduced sample (shaded in grey). Left (right) panels correspond to year 2007 (2008). Top panels show the bi-Gaussian fittings (in red) with the corresponding single curves in blue and green. The red curve in bottom panels corresponds to the single Gaussian fittings.

Figure 11 shows the scatter plot of  $p_1$  and  $p_2$  versus the SSN for the reduced sample. The values of the Pearson correlation coefficient  $r(p_1, SSN) = 0.41$  and  $r(p_2, SSN) = 0.76$  are similar to those obtained before, showing a strong (weak) linear relationship of  $p_2$  ( $p_1$ ) with the solar cycle, but this time including year 2003 in the sample. Indeed, the linear relationship of  $p_1$  with the solar cycle is so weak that the interval centered in the weighted average with an uncertainty of  $\pm\sigma_w$  ( $p_1 = 9.6 \pm 0.3$ ) includes the regression line. Note also that the y-intercept of the linear regression between  $p_2$  and the SSN ( $p_2 = 9.8$ ) is included in the interval  $p_1 = 9.6 \pm 0.3$ . This result allows us to recognize that the type of wind labeled with the subindex 1 corresponds to the slow wind. Indeed, in the case of no sunspot (i.e. at the y-intercept) no relevant coronal holes out of the poles are foreseen (as expected in the solar minimum) and therefore the only type of wind will be the slow one. This case would be similar to year 2009, where a single Gaussian function is enough to describe the whole bulk solar wind. Thus, for  $SSN=0$ , both single Gaussian curves are expected to coincide.



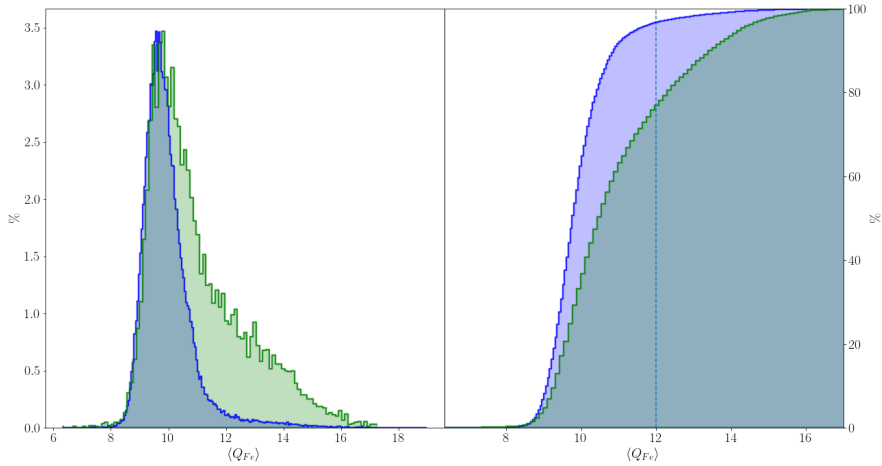
**Figure 11.** Scatter plots of  $p_1$  (left) and  $p_2$  (right) vs. SSN. The blue shadowed area represents the weighted average of  $p_1 \pm \sigma_w$ . The orange solid lines correspond to the linear regression and the orange dashed lines in the right panel are the 99% confidence intervals.

#### 4. Are ICMEs the Outliers?

The whole sample of  $\langle Q_{Fe} \rangle$  data from SWICS/ACE has been split into three sets after the workflow in Figure 6: bulk solar wind (or reduced sample), ICMEs previously identified, and outliers. The set of outliers includes all data where  $\langle Q_{Fe} \rangle > 12$ , which were not previously identified as ICME material. We ask now, which type of solar wind are the outliers? Where do they come from? Considering the physical processes happening in the solar atmosphere, they cannot be part of the slow solar wind nor of the fast wind coming from coronal holes. Indeed, Lepri et al. (2001) show that the  $\langle Q_{Fe} \rangle$  of the bulk solar wind are typically around 9 to 11.

The right panel in Figure 12 compares the cumulative distribution function of  $\langle Q_{Fe} \rangle$  for the whole ACE data set (in blue) and for the ICMEs from the catalogs described above (in green). It can be appreciated that more than 95% (exactly 96.7%) of the values of the  $\langle Q_{Fe} \rangle$  are below 12 for the whole solar wind sample, reinforcing the result by Lepri (2004). This value drops to 76.9% for the data sample from the ICME catalogs and, although just 23.1% of the values of  $\langle Q_{Fe} \rangle$  from the ICMEs are larger than 12, the greater-than-12 part of the distribution function is very distinct to that of the whole sample (left panel Figure 12). The less-than-12 part of the distribution function of the ICME values may be related to ICMEs with no large flaring activity (avoiding large ionization states of Fe) and/or to ICMEs containing also normal solar wind. Indeed, Richardson and Cane (2005) noted that different signatures of an ICME may not occur exactly concurrently. Being aware of this problem, Jian et al. (2006) included in their ICME intervals the shock (if it occurs), sheath pile-up region, and the ejecta.

Certainly, high charge states arise because high temperatures in the solar corona, associated with the initiation of CMEs, ionize the material ejected into the solar wind.



**Figure 12.** (Left figure) Distribution function of  $\langle Q_{Fe} \rangle$  for the whole ACE data set (blue) and for the ICMEs from the catalogs (green). The distribution function of the ICMEs has been rescaled for comparison. (Right figure) Cumulative distribution function of  $\langle Q_{Fe} \rangle$  for the whole ACE data set (blue) and for the ICMEs from the catalogs (green).

Thus, large deviations from typical values of  $\langle Q_{Fe} \rangle$  are expected to be associated with ICMEs. Therefore, the goal of this section is to analyze the solar wind parameters during outliers with the aim of identifying any signature of ICME material, other than enhanced  $\langle Q_{Fe} \rangle$ . For this purpose, we have arranged the outliers as a list of events, considering an event when  $\langle Q_{Fe} \rangle > 12$  during at least 10 hours consecutively. The result includes 27 events, which are listed in Table 1. The times listed in Table 1 are accurate to within 2 hours (the time resolution of the data set).

The events have been classified into two groups (see column 3 in Table 1): 'Extended' and 'New'. Those cataloged as 'Extended' are events where an extension of the boundaries of an already identified ICME will include the outlier event. When there is no identified ICME close to the outlier event, it is labeled as 'New'.

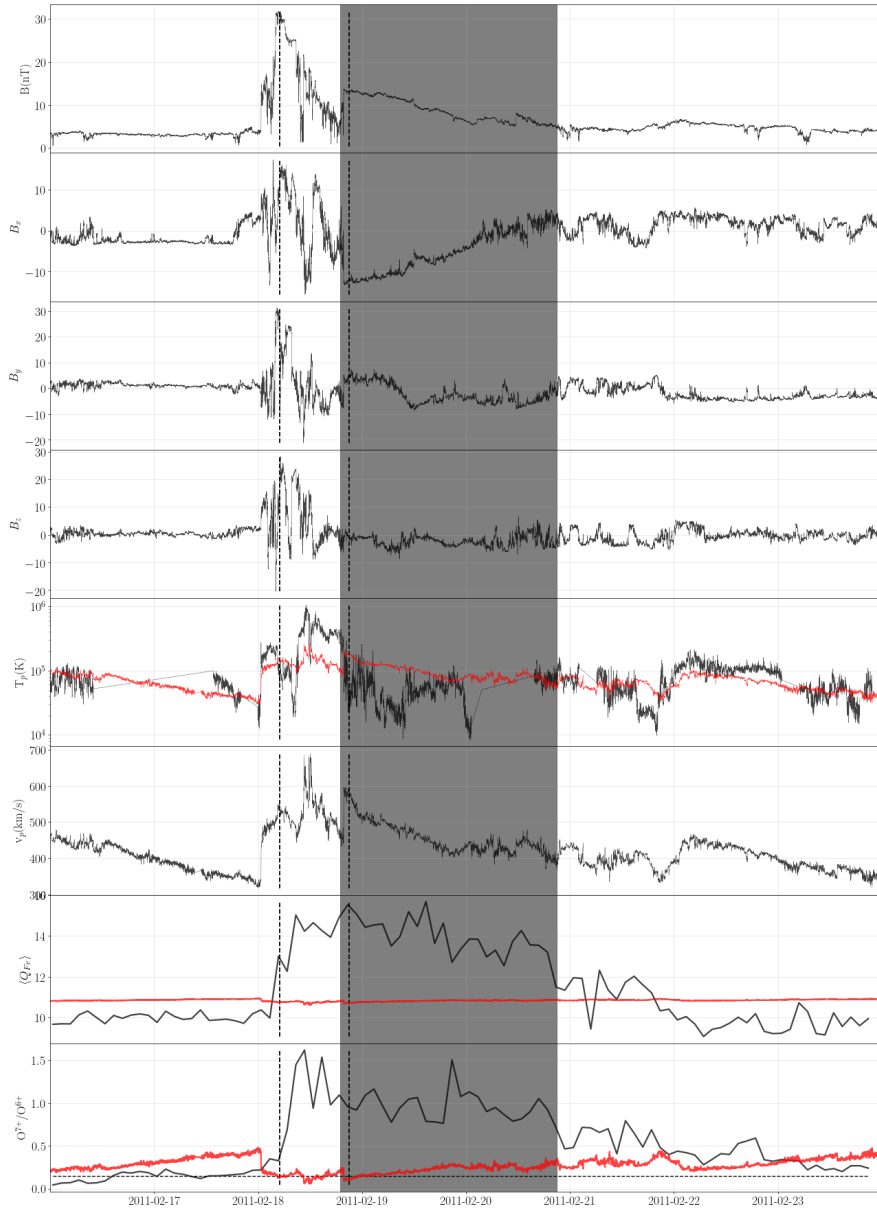
Figure 13 shows the solar wind parameters measured by ACE during the event on 18 February 2011, as an example of an 'Extended' event. From top to bottom Figure 13 shows the magnetic field strength ( $B$ ) and its GSM components ( $B_x$ ,  $B_y$ ,  $B_z$ ) from the Magnetic Field Experiment instrument (MAG/ACE); the proton temperature ( $T_p$ ), and the proton speed ( $v_p$ ) from the Solar Wind Electron, Proton, and Alpha Monitor instrument (SWEPAM/ACE), and the  $\langle Q_{Fe} \rangle$ , and  $O^{7+}/O^{6+}$  ratio from SWICS/ACE, from 16 to 23 February 2011. Superimposed on the observed values of  $T_p$ ,  $\langle Q_{Fe} \rangle$ , and  $O^{7+}/O^{6+}$  ratio, as red solid lines, are the expected values according to Richardson and Cane (1995) and Richardson and Cane (2004). The horizontal black dashed line in the bottom panel indicates the reference  $O^{7+}/O^{6+} = 0.145$ . Zhao, Zurbuchen, and Fisk (2009) and Heidrich-Meisner et al. (2016) state that values below this threshold correspond to fast streams coming from coronal holes.

The shadowed area in Figure 13 corresponds to the ICME interval as identified in the catalogs previously detailed. As there is not an agreement on the boundaries of

**Table 1.** List of outlier events

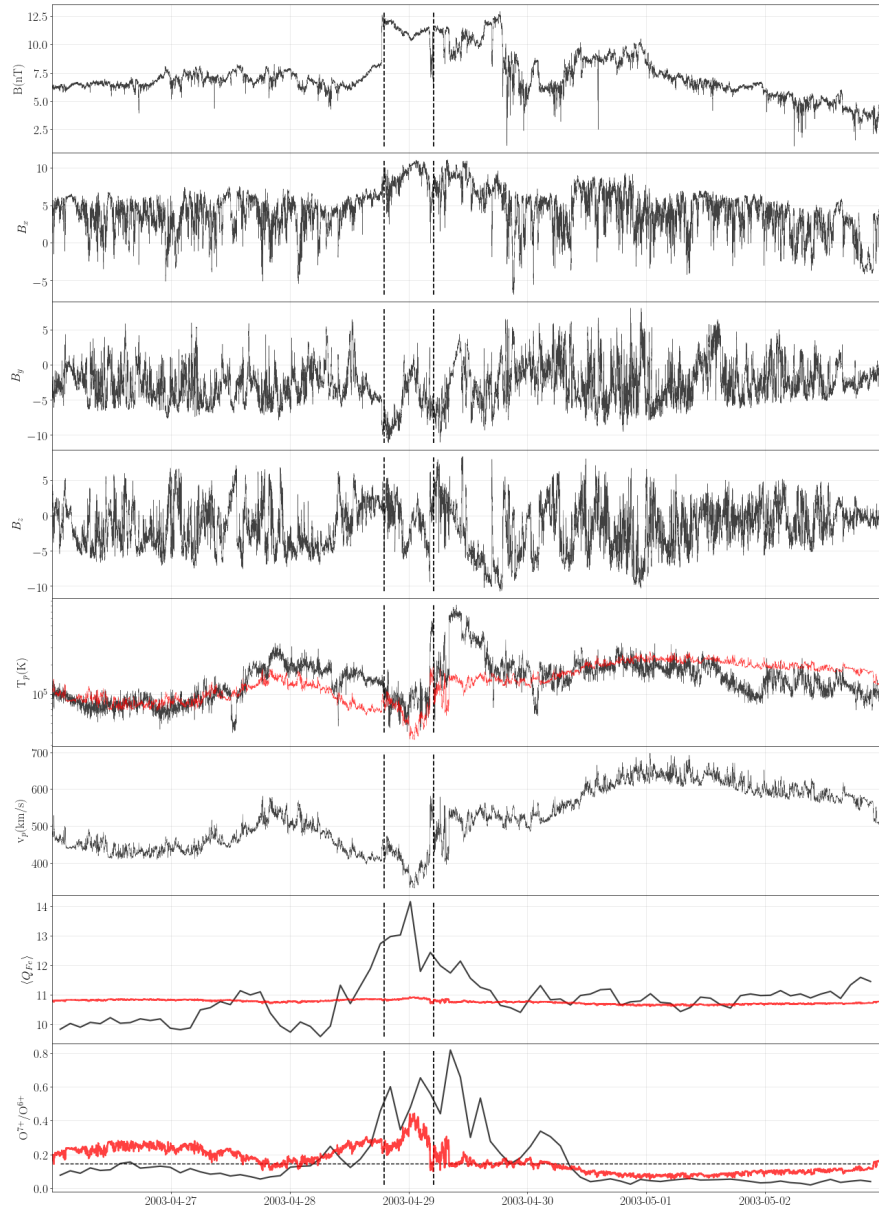
$\langle Q_{Fe} \rangle > 12$ Start year month day UT	$\langle Q_{Fe} \rangle > 12$ End year month day UT	Type of event
1999 01 09 18h	1999 01 10 04h	New
1999 11 15 00h	1999 11 15 16h	Extended
2000 03 11 05h	2000 03 11 13h	New
2000 07 17 09h	2000 07 17 23h	Extended
2000 12 24 06h	2000 12 24 14h	Extended
2002 06 27 12h	2002 06 27 22h	New
2002 09 23 06h	2002 09 23 16h	Extended
2002 09 26 06h	2002 09 26 14h	New
2003 04 28 19h	2003 04 29 05h	New
2003 07 08 17h	2003 07 09 03h	Extended
2003 07 09 11h	2003 07 09 19h	Extended
2003 07 10 23h	2003 07 11 09h	Extended
2003 10 30 11h	2003 10 31 03h	Extended
2003 11 02 00h	2003 11 02 08h	Extended
2005 01 20 03h	2005 01 20 13h	Extended
2005 07 12 15h	2005 07 13 01h	Extended
2005 12 07 10h	2005 12 09 08h	New
2010 02 15 01h	2015 02 15 11h	New
2011 02 18 05h	2011 02 18 21h	Extended
2011 08 06 08h	2011 08 06 20h	Extended
2012 06 18 02h	2012 06 18 10h	Extended
2013 05 22 06h	2013 05 22 14h	New
2013 10 27 19h	2013 10 28 03h	New
2015 01 29 14h	2015 01 30 06h	New
2015 02 04 20h	2015 02 05 12h	New
2015 03 05 04h	2015 03 05 16h	New
2016 07 22 15h	2016 07 24 03h	Extended

this event (as in many others), we have drawn the shadowed area so that it covers the greater interval considering the boundaries of different catalogs. In Figure 13, the start date of the shadowed area comes from the Richardson and Cane catalog and the end date from the Wind catalog. The dashed vertical lines indicate the boundaries of the region where  $\langle Q_{Fe} \rangle > 12$ . Values of  $\langle Q_{Fe} \rangle$  over 14 reached in the interval of the outlier are difficult to explain if this solar wind does not come from a CME. The larger values of the  $O^{7+}/O^{6+}$  ratio relative to the expected one also support the hypothesis of CME material in this interval. Between the front boundary of the outlier (first dashed line) and the beginning of the shadowed area, the magnetic field strength reached more than 30 nT, about more than twice the one in the shadowed area. This high value in the magnetic field strength may be related to the interaction between a halo CME launched on 15 February 2011 at 02:24UT and a previous slower halo CME first appearing on the second telescope of the Large Angle and Spectrometric Coronagraph (C2/LASCO) on 14 February 2011 at 18:24UT. In this scenario, the outlier interval corresponds to



**Figure 13.** Solar wind parameters from 16 to 23 February 2011. From top to bottom: magnetic field strength ( $B$ ), magnetic field components ( $B_x$ ,  $B_y$ ,  $B_z$ ), proton temperature ( $T_p$ ), proton speed ( $v_p$ ), average iron charge state  $\langle Q_{Fe} \rangle$  and oxygen ratio  $O^{7+}/O^{6+}$





**Figure 14.** Solar wind parameters from 26 April to 2 May 2003. From top to bottom: magnetic field strength ( $B$ ), magnetic field components ( $B_x$ ,  $B_y$ ,  $B_z$ ), proton temperature ( $T_p$ ), proton speed ( $v_p$ ), average iron charge state  $\langle Q_{Fe} \rangle$  and oxygen ratio  $O^{7+}/O^{6+}$

the first ICME which is compressed by the second one, appearing in the shadowed area. As in the event analyzed by Cid et al. (2016), during this event the boundary identification from magnetic signatures does not agree with those from other signatures. This discrepancy can be explained as due to the interaction of transients, which modifies the magnetic topology of ICMEs.

Detailed analysis of the events in Table 1 guide us to conclude that 14 additional events present similar features to those described above during the event on 18 February 2011, where an extension of the ICME boundaries allow to consider the outliers as ICME material. But what about the other 12 events (almost half of the sample)? Values of  $\langle Q_{Fe} \rangle$  over 12 for more than 10 hours are difficult to explain if the solar wind transient is not an ICME. In some cases the ICME might have been missed in the catalogs due to data gap in any relevant magnitude used during the identification process. For example, a data gap in temperature of about one day appears during the event on 11 March 2000, but there are several events without data gaps. The event on 28–29 April 2003 is an example of the ‘New’ events without data gap. Solar wind parameters during that event appear in Figure 14 with the same format as in Figure 13. The interval of the outlier coincides with a region where the magnetic field is enhanced up to more than 10 nT and the proton temperature is below the expected one, according to Richardson and Cane (1995). The outlier interval is surrounded by solar wind with high speed (500–600 km/s) and with both magnetic field vector and velocity highly fluctuating, as expected from the fast wind coming from a coronal hole. Indeed, the  $O^{7+}/O^{6+}$  ratio shows values below 0.145, reinforcing the signatures of fast streams. In this scenario, the large discontinuities in the magnetic field strength at the boundaries of the outlier appear as signatures of the interaction of the surrounding fast wind with an ICME (identified as a ‘New’ outlier event). Moreover, before and after the outlier, the value of  $\langle Q_{Fe} \rangle$  is around 10 and 11, respectively, which is compatible with the type of wind labeled with subindex 2 (see Figure 7.)

## 5. Summary and Conclusions

Here we show that the bi-Gaussian function reproduces the empirical distribution function of the  $\langle Q_{Fe} \rangle$  of the bulk solar wind at 1 AU. This bimodal wind presents two components which spread around  $\langle Q_{Fe} \rangle = 9.6$  and 10.3, with the last one strongly dependent on the solar cycle. These two components are supposed to be associated with slow and fast wind. Nevertheless, Figure 5 and Figure 11 demonstrate that  $p_2$  is related with the solar cycle. The relationship of the number of active regions and the SSN is undeniable. Thus, this result supports the results from Neugebauer et al. (2002) identifying two solar sources of the fast solar wind: the coronal holes and the active regions.

Our results are obtained from the analysis of the whole data set of  $\langle Q_{Fe} \rangle$  from SWICS/ACE from 1998 to 2017, after separating the bulk solar wind from the transients. The transients include, not only the events previously identified as ICMEs in different sources, but also those intervals where  $\langle Q_{Fe} \rangle > 12$ , which we labeled as outliers. The analysis of the whole data sample shows that the threshold set by Lepri (2004) in  $\langle Q_{Fe} \rangle = 12$  for the ICMEs, corresponds to a deviation of  $3\sigma$  from the position of the peak of the second Gaussian distribution, supporting the establishment of this

threshold as robust for ICMEs, as ICMEs are the unique outlier of the bulk solar wind known to date.

We also provide a catalog of 27 outliers where the condition  $\langle Q_{Fe} \rangle > 12$  is maintained at least for 10 hours. From the analysis of the different solar wind parameters around the events of the catalogue, we report that half of the events can be considered as part of an already identified ICME or a group of ICMEs. We identify new ICME events in the other half of the catalog. These events may be missing in the existing catalogs because of a data gap in one of the solar wind parameters commonly used to identify ICMEs or due to an anomalous behavior of some of the parameters due to interaction of the CME material with the surrounding solar wind or even with other ICMEs. From our results we strongly support that  $\langle Q_{Fe} \rangle > 12$  is a sufficient signature to identify ICMEs in the solar wind and the most convenient signature to identify its boundaries. Future work will be dedicated to understand whether the lower-than-12 values of  $\langle Q_{Fe} \rangle$  in ICMEs are associated with their solar origin (i.e. non large flaring) or to the boundary identifications.

**Acknowledgements** This work was supported by the MINECO project AYA2016-80881-P (including FEDER funds). We thank the SWICS, SWEPAM and MAG instruments teams and the ACE Science Center for providing the ACE data. We acknowledge WDC-SILSO, Royal Observatory of Belgium, Brussels for providing the sunspot number. We also acknowledge the information from the CME catalog generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. The authors want to thank an anonymous reviewer for the useful comments. Disclosure of Potential Conflicts of Interest: The authors declare that there are no conflicts of interest.

## References

- Bale, S.D., Badman, S.T., Bonnell, J.W., Bowen, T.A., Burgess, D., Case, A.W., Cattell, C.A., Chandran, B.D.G., Chaston, C.C., Chen, C.H.K., et al.: 2019, Highly structured slow solar wind emerging from an equatorial coronal hole. *Nature*, 1.
- Banaszkiewicz, M., Czechowski, A., Axford, W.I., McKenzie, J.F., Sukhorukova, G.V.: 1997, The Fast Solar Wind and its Source Region. In: Wilson, A. (ed.) *Correlated Phenomena at the Sun, in the Heliosphere and in Geospace, ESA Special Publication* **415**, 17. ADS.
- Burlaga, L.F., King, J.H.: 1979, Intense interplanetary magnetic fields observed by geocentric spacecraft during 1963-1975. *Journal of Geophysical Research* **84**(A11), 6633. DOI. ADS.
- Cid, C., Palacios, J., Saiz, E., Guerrero, A.: 2016, Redefining the boundaries of interplanetary coronal mass ejections from observations at the ecliptic plane. *Astrophysical Journal* **828**(1), 11. DOI.
- Cranmer, S.R., Gibson, S.E., Riley, P.: 2017, Origins of the Ambient Solar Wind: Implications for Space Weather. *Space Science Reviews* **212**(3-4), 1345. DOI. ADS.
- Gloeckler, G., Cain, J., Ipavich, F.M., Tums, E.O., Bedini, P., Fisk, L.A., Zurbuchen, T.H., Bochsler, P., Fischer, J., Wimmer-Schweingruber, R.F., Geiss, J., Kallenbach, R.: 1998, Investigation of the composition of solar and interstellar matter using solar wind and pickup ion measurements with swics and swims on the ace spacecraft. *Space Science Reviews* **86**(1), 497. DOI.
- Gringauz, K.I.: 1961, Some Results of Experiments in Interplanetary Space by Means of Charged Particle Traps on Soviet Space Probes. In: *Space Research II*, 539. ADS.
- Gringauz, K.I., Bezrukh, V.V., Musatov, L.S.: 1967, Solar-Wind Observations with the Venus 3 Probe. *Cosmic Research* **5**, 216. ADS.

- Gringauz, K.I., Bezrokhikh, V.V., Ozerov, V.D., Rybchinskii, R.E.: 1960, A Study of the Interplanetary Ionized Gas, High-Energy Electrons and Corpuscular Radiation from the Sun by Means of the Three-Electrode Trap for Charged Particles on the Second Soviet Cosmic Rocket. *Soviet Physics Doklady* **5**, 361. ADS.
- Heidrich-Meisner, V., Peleikis, T., Kruse, M., Berger, L., Wimmer-Schweingruber, R.: 2016, Observations of high and low Fe charge states in individual solar wind streams with coronal-hole origin. *Astronomy & Astrophysics* **593**, A70. DOI. ADS.
- Hundhausen, A.J.: 1972, Coronal Expansion and Solar Wind. *Physics and Chemistry in Space* **5**. DOI. ADS.
- Jian, L.K., Russell, C.T., Luhmann, J.G.: 2011, Comparing Solar Minimum 23/24 with Historical Solar Wind Records at 1 AU. *Solar Physics* **274**(1-2), 321. DOI. ADS.
- Jian, L., Russell, C.T., Luhmann, J.G., Skoug, R.M.: 2006, Properties of Interplanetary Coronal Mass Ejections at One AU During 1995-2004. *Solar Physics* **239**(1-2), 393. DOI. ADS.
- Larrodera, C., Cid, C.: 2020, Bimodal distribution of the solar wind at 1 au. *Astronomy & Astrophysics* **635**, A44.
- Lepri, S.T.: 2004, Iron charge state distributions as an indicator of hot ICMEs: Possible sources and temporal and spatial variations during solar maximum. *Journal of Geophysical Research* **109**(A1). DOI.
- Lepri, S.T., Zurbuchen, T.H., Fisk, L.A., Richardson, I.G., Cane, H.V., Gloeckler, G.: 2001, Iron charge distribution as an identifier of interplanetary coronal mass ejections. *Journal of Geophysical Research* **106**(A12), 29231. DOI. ADS.
- Li, K.J., Zhanng, J., Feng, W.: 2016, A Statistical Analysis of 50 Years of Daily Solar Wind Velocity Data. *Astrophysical Journal* **151**, 128. DOI. ADS.
- Lionello, R., Riley, P., Linker, J.A., Mikić, Z.: 2005, The Effects of Differential Rotation on the Magnetic Structure of the Solar Corona: Magnetohydrodynamic Simulations. *Astrophysical Journal* **625**(1), 463. DOI. ADS.
- Neugebauer, M., Snyder, C.W.: 1966, Mariner 2 Observations of the Solar Wind, 1, Average Properties. *Journal of Geophysical Research* **71**, 4469. DOI. ADS.
- Neugebauer, M., Liewer, P.C., Smith, E.J., Skoug, R.M., Zurbuchen, T.H.: 2002, Sources of the solar wind at solar activity maximum. *Journal of Geophysical Research* **107**(A12), 1488. DOI. ADS.
- Richardson, I., Cane, H.: 2005, Survey of Interplanetary Coronal Mass Ejections in the Near-Earth Solar Wind During 1996 – 2005. In: *AGU Spring Meeting Abstracts* **2005**, SH43A. ADS.
- Richardson, I.G., Cane, H.V.: 1995, Regions of abnormally low proton temperature in the solar wind (1965-1991) and their association with ejecta. *Journal of Geophysical Research* **100**(A12), 23397. DOI. ADS.
- Richardson, I.G., Cane, H.V.: 2004, Identification of interplanetary coronal mass ejections at 1 AU using multiple solar wind plasma composition anomalies. *Journal of Geophysical Research* **109**(A9). DOI.
- Rossi, B.: 1991, The interplanetary plasma. *Annual Review of Astronomy and Astrophysics* **29**, 1. DOI. ADS.
- Schwenn, R.: 2006a, Solar Wind Sources and Their Variations Over the Solar Cycle. *Space Science Reviews* **124**, 51. DOI. ADS.
- Schwenn, R.: 2006b, Space Weather: The Solar Perspective. *Living Reviews in Solar Physics* **3**(1), 2. DOI. ADS.
- Viall, N.M., Borovsky, J.E.: 2020, Nine outstanding questions of solar wind physics. *Journal of Geophysical Research: Space Physics*. DOI.
- Vörös, Z., Leitner, M., Narita, Y., Consolini, G., Kovács, P., Tóth, A., Lichtenberger, J.: 2015, Probability density functions for the variable solar wind near the solar cycle minimum. *Journal of Geophysical Research* **120**(8), 6152. DOI. ADS.
- Zhao, L., Zurbuchen, T.H., Fisk, L.A.: 2009, Global distribution of the solar wind during solar cycle 23: ACE observations. *Geophysical Research Letters* **36**(14), L14104. DOI. ADS.