

**Special Section:**

Cluster 20th anniversary: results from the first 3D mission

**Key Points:**

- We use Cluster multipoint measurements to investigate the different magnetosheath plasma environments according to the upstream conditions
- Local magnetosheath measurements can be used to characterize and classify the magnetosheath plasma with respect to upstream conditions
- We suggest future extensions to the methodology

**Correspondence to:**

T. Karlsson,  
tomas.karlsson@ee.kth.se

**Citation:**

Karlsson, T., Raptis, S., Trollvik, H., & Nilsson, H. (2021). Classifying the magnetosheath behind the quasi-parallel and quasi-perpendicular bow shock by local measurements. *Journal of Geophysical Research: Space Physics*, 126, e2021JA029269. <https://doi.org/10.1029/2021JA029269>

Received 22 FEB 2021

Accepted 30 AUG 2021

## Classifying the Magnetosheath Behind the Quasi-Parallel and Quasi-Perpendicular Bow Shock by Local Measurements

T. Karlsson<sup>1</sup> , S. Raptis<sup>1</sup> , H. Trollvik<sup>1</sup>, and H. Nilsson<sup>2</sup> 

<sup>1</sup>Division of Space and Plasma Physics, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden, <sup>2</sup>Swedish Institute of Space Physics, Kiruna, Sweden

**Abstract** We investigate and evaluate the possibility of using local magnetosheath measurements to classify the plasma according to upstream conditions. In order to do this, we use simultaneous measurements from the Cluster spacecraft from time intervals when one of them is located in the solar wind, and the other in the magnetosheath. In particular, we study the classification of the magnetosheath plasma into the classes quasi-parallel versus quasi-perpendicular and foreshock/no foreshock (referring to the geometry of the upstream bow shock). We evaluate this method based on the magnetosheath measurements of the high-energy ion energy flux, magnetic field standard deviation, and ion temperature anisotropy. We find that the method is promising and useful, in that it eliminates the uncertainties associated with propagating upstream measurements made far from the bow shock. Finally, we discuss some possible extensions of the methodology to be investigated in the future.

### 1. Introduction

The terrestrial bow shock is a thin region where the solar wind transitions from a supersonic to a subsonic velocity as it interacts with the geomagnetic field. As the solar wind plasma crosses the bow shock, it is heated and compressed, forming the magnetosheath. Many of the properties of the bow shock depend on the orientation of the interplanetary magnetic field (IMF) with respect to the bow shock. This is typically expressed in terms of the angle  $\theta_{Bn}$ , the angle between the IMF and the shock normal (e.g., Balogh et al., 2005, and references therein). For  $\theta_{Bn} < 45^\circ$ , the shock is denoted as quasi-parallel, while a shock with  $\theta_{Bn} > 45^\circ$  is called a quasi-perpendicular shock.

Closely related to the classification of the bow shock into quasi-parallel and quasi-perpendicular configurations is the concept of the foreshock. The foreshock is defined as the region behind the last interplanetary field line that connects to the bow shock, "the tangent field line," which defines the foreshock boundary. Some distance away from the tangent field line, toward the quasi-parallel region, a field-aligned ion beam (FAB) is typically observed in a relatively thin region. This defines the boundary of the ion foreshock. Between the tangent field line and the FAB, backstreaming electrons are observed in the electron foreshock. The FAB connects to the bow shock at a position where  $\theta_{Bn} = 40 - 70^\circ$  (e.g., Wilson, 2016), where the exact position may depend on the IMF direction (Le & Russell, 1992). The ion foreshock therefore typically connects to the whole quasi-parallel part of the bow shock, but also extends into the quasi-perpendicular region.

Deeper inside the foreshock, where  $\theta_{Bn}$  decreases, other types of ion populations are present: intermediate, gyrating, and diffuse ions. These ion distributions are often organized according to the distance from the FAB and tangential field line, with the diffuse ions found deepest inside the foreshock (e.g., Eastwood et al., 2005; Kis et al., 2007; Tsurutani & Rodriguez, 1981; Wilson, 2016). Of these ion distributions, the diffuse ions typically have the highest energies, sometimes over 150 keV (Eastwood et al., 2005). The details of how these ion populations are generated and their relation to each other are not fully known (e.g., Eastwood et al., 2005).

Another important feature of the foreshock is the presence of various types of ultralow frequency (ULF) fluctuations (Eastwood et al., 2005; Wilson, 2016), generated by the interaction of the suprathermal ion distributions with the solar wind (Gary, 1991). The ULF fluctuations are usually located some distance away

© 2021. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

from the FAB foreshock boundary and are rarely observed together with the FAB (Eastwood et al., 2005). Deeper inside the foreshock, the waves further interact with the suprathermal ion distributions resulting in more broken up, nonlinear structures such as shocklets and SLAMS (Short Large-Amplitude Magnetic Structures) (e.g., Battarbee et al., 2020; Eastwood et al., 2005).

The different upstream environments, which depend on the angle  $\theta_{Bn}$  at which the solar wind field lines connect to the bow shock, are likely to be transmitted to the downstream magnetosheath region. Fuselier (1994) introduced the terms "quasi-parallel magnetosheath" and "quasi-perpendicular magnetosheath" for the magnetosheath regions located on the streamline connecting to the quasi-parallel and quasi-perpendicular bow shocks, respectively, and suggested that the high-energy ion populations in the magnetosheath were produced at the quasi-parallel bow shock and transmitted downstream (Fuselier, 1994; Fuselier et al., 1991). In an earlier study, Crooker et al. (1981) showed that energetic ions (>3 keV) were present in the magnetosheath on streamlines connecting to bow shock regions where  $\theta_{Bn} < 60^\circ$ . Chang et al. (2000) have also shown that there was a clear correlation between magnetosheath proton flux and the IMF cone angle (which was used as a proxy for  $\theta_{Bn}$ ) during one day of observations. Also other properties of the magnetosheath seem to be correlated with the upstream conditions: Luhmann et al. (1986) showed that the magnetic field fluctuation level was correlated with the IMF direction and concluded that the quasi-parallel bow shock was a strong source of dayside magnetosheath fluctuations, interpreting them as upstream magnetic field fluctuations convecting across the bow shock. Finally, Dimmock et al. (2015) reported on a strong local-time asymmetry in the ion temperature anisotropy in the magnetosheath, showing a strong IMF dependence possibly related to processes behind the quasi-parallel and quasi-perpendicular bow shocks.

The interest of the authors of this paper in the connection between the upstream and downstream plasma properties has mainly been triggered by the connection of magnetosheath jets (localized, intermittent regions of elevated dynamic pressure; e.g., Archer & Horbury, 2013; Hietala & Plaschke, 2013; Karlsson et al., 2018; Plaschke et al., 2013, 2017) to certain upstream conditions. Early results have shown that magnetosheath jets are most often observed during the conditions of radial IMF, which correspond to a quasi-parallel bow shock on the day side, where most of the statistics on magnetosheath jets have been collected (Archer & Horbury, 2013; Plaschke et al., 2013). However, it has been shown that magnetosheath jets can also be found in the quasi-perpendicular magnetosheath (Goncharov et al., 2020; Raptis, Karlsson, et al., 2020; Vuorinen et al., 2019).

In most of these studies, the upstream conditions have been characterized by a determination of the angle  $\theta_{Bn}$  in the upstream region associated with the magnetosheath jets, based on measurements in the solar wind. This has several drawbacks; the upstream data are not always available, the temporal resolution may be low, and the propagation effects from the solar wind measurements to relate them to the magnetosheath measurements introduce large uncertainties. This is particularly true for measurements close to the flanks where misclassification can occur regularly (Raptis, Aminalragia-Giamini, et al., 2020). Instead Raptis, Karlsson, et al. (2020) based the classification of the magnetosheath jets on the properties of the local MMS (Burch et al., 2016; Fuselier et al., 2016) magnetosheath measurements, making use of the connection to the upstream properties described above. Raptis, Karlsson, et al. (2020) used a combination of the magnetosheath levels of energetic ion fluxes, magnetic field fluctuations, and ion temperature anisotropy to classify magnetosheath jets as located in the quasi-parallel or quasi-perpendicular magnetosheath (or the boundary between them). They did provide some verification of the methodology by comparing to upstream IMF measurements. However, there is a need to better verify this method and investigate how well it works. This is the purpose of this paper, which will present a first step toward characterizing the local magnetosheath according to the upstream conditions, without actual upstream measurements. Such a classification can of course also be useful for many other types of studies, e.g., constructing extended databases to investigate the sources of magnetosheath fluctuations and turbulence.

In order to investigate the classification methodology, we will make use of measurements where one of the Cluster spacecraft is located upstream of the bow shock and another in the magnetosheath. The relative small spacecraft separations (1–2  $R_E$ ) allow us to correlate the downstream conditions with the upstream ones, without large uncertainties due to the propagation of upstream measurements, as discussed above.

## 2. Data and Methods

For this study, we use data from the Cluster spacecraft, which move in formation with satellite separations varying between 200 and several tens of thousands of kilometers during the mission (e.g., Escoubet et al., 1997; ESA, 2020). During the times of larger separations, sometimes, one spacecraft can be located in the solar wind, while another is in the magnetosheath for rather extended times. We will use such time intervals to investigate the connections between magnetosheath properties and solar wind conditions.

In order to do this, we use data from the FluxGate Magnetometer (FGM) (Balogh et al., 2001) and the Cluster Ion Spectrometer (CIS) (Dandouras & Barthe, 2012; Réme et al., 1997). For the magnetic field data, we use the C[n]\_CP\_FGM\_SPIN data set, which gives the full magnetic field vector with spin resolution ( $\sim 4$  s) in the GSE coordinate system (Carr et al., 2013). For the ion data, we use spacecraft 3 (C3) as the magnetosheath monitor with data from the Hot Ion Analyzer (HIA) (data product C3\_CP\_CIS-HIA\_HS\_SWIONS\_PEF). For the solar wind monitor for the first part of this study, we use C4 with its Composition and Distribution Function (CODIF) analyzer data set C4\_CP\_CIS-CODIF\_HS\_H1\_PEF, which contains the proton energy fluxes. For the second part (Section 3.3), we use C1 as the solar wind monitor (data set C1\_CP\_CIS-HIA\_HS\_SWIONS\_PEF). We also use the data set C[n]\_CP\_AUX\_POSGSE\_1M for spacecraft positions. All data have been downloaded from the open Cluster Science Archive (Laakso et al., 2010). The preference for C3 for the magnetosheath monitor is due to the fact that for the time intervals studied, this data set usually had the highest time and energy resolution for the ion measurements.

In order to find suitable time intervals, where one spacecraft is situated in the solar wind and one in the magnetosheath, we have manually inspected data from January to April, for the years 2003, 2006, and 2009. For these times, the satellite separations were large enough, and the CIS instruments were operational on S/C 1, 3, and 4. To select suitable time intervals for the first part of this study, we used the following criteria for the solar wind and magnetosheath data, which should all be fulfilled simultaneously:

1. Only solar wind data within 30 min of a bow shock crossing are considered.
2. When the solar wind time interval is bounded by a region containing the magnetosheath plasma, the latter region should be at least 10-min long.
3. The solar wind time interval should be at least 10-min long.
4. For the whole solar wind interval considered, S/C 3 should be located in the magnetosheath.
5. Magnetosheath data within 10 min of a magnetopause crossing are not considered.
6. A magnetic field structure, clearly identifiable in both the solar wind and magnetosheath data, should exist in order to establish a time lag between the solar wind and magnetosheath data. See below for further description of the lag methodology.

Determining magnetosheath (MSH), solar wind (SW), and magnetosphere regions is done manually, with the help of ion spectrograms, ion flow velocities, and magnetic field data.

The above criteria are designed to find the time intervals of relatively stable conditions, where determination of, e.g.,  $\theta_{Bn}$ , the angle between the solar wind magnetic field and the bow shock, can be calculated with high confidence. With the application of these criteria, the resulting data intervals are shown in Table 1.

An example of a time interval used is shown in Figure 1. As can be seen from panel (a), S/C 3 enters the magnetosheath at around 19:22 UTC and remains there for the rest of the time interval shown. S/C 4, meanwhile, is clearly located in the solar wind, until about 19:55 UTC (but see below regarding time shifts). For the time interval chosen for this study (19:23–19:53 UTC), a number of quantities have been calculated in the following way. (Actually, the quantities for S/C 3 were also calculated outside of this time interval, as can be seen in the plot, but values outside of the time interval were disregarded.)

The calculation of some of these quantities is further described in the text. The S/C 4 data have been shifted in time, as described below.

For S/C 3, Figure 1e shows a running average of the magnetic field constructed by calculating the standard deviation  $\sigma$  of each magnetic field component for a 30-s window. The total standard deviation is then calculated as

$$\sigma(\mathbf{B}) = \sqrt{\sigma(B_x)^2 + \sigma(B_y)^2 + \sigma(B_z)^2} \quad (1)$$

**Table 1**  
*Data Used for Correlation Between  $\theta_{Bn}$  and Magnetosheath Properties*

Date	Start time	Stop time	SW S/C	MSh S/C	Bow shock time	Lag (s)	$X_{GSE}(R_E)$	$Y_{GSE}(R_E)$	$Z_{GSE}(R_E)$
2003-03-03	14:23	14:36	C4	C3	14:21:55	-23	9.0	3.3	7.4
2006-02-10	05:34	06:04	C4	C3	06:05:20	-71	9.0	-1.2	-10.9
2006-02-19	19:23	19:53	C4	C3	19:54:30	-36	7.3	-3.2	-10.7
2006-04-05	22:30	22:50	C4	C3	22:51:50	-16	5.0	-8.5	-11.1
2006-04-05	23:12:30	23:28	C4	C3	23:11:10	-16	4.3	-8.2	-11.0
2006-04-22	10:05	10:18	C4	C3	10:04:00	0	5.5	-11.9	-11.0
2009-03-17	05:50	06:10	C4	C3	05:38:30	-42	14.8	2.2	-3.1
2009-03-24	07:35	08:05	C4	C3	07:33:40	-89	13.8	1.0	-2.4
2009-03-29	03:21	03:51	C4	C3	03:19:45	-120	15.0	-0.9	-3.3

*Note.* All times are given in UTC. The spacecraft in the solar wind (SW) and magnetosheath (MSh) are indicated by their standard numbers. The start and stop times refer to the solar wind spacecraft. The indicated bow shock time is used for the bow shock model required to determine  $\theta_{Bn}$ . The positions are given for S/C 1, for the middle of the time interval.

where  $B_x$ , etc., are the GSE components. Panel (f) shows the ion temperature anisotropy, defined as

$$Q = \frac{T_{i,\perp}}{T_{i,\parallel}} - 1 \quad (2)$$

where  $T_{i,\perp}$  and  $T_{i,\parallel}$  are the ion perpendicular and parallel temperatures, respectively.

Panel (g) shows the energy flux of high-energy ions, calculated by integrating the energy flux for the four highest energy channels of HIA on S/C 3:

$$F_{high} = \sum_{n=1}^4 F(E_i) \Delta E_i \quad (3)$$

where  $E_i, i = 1, 2, . . . 4$  are the energies of the highest energy channels,  $F(E_i)$  is the differential energy flux of that energy channel, and  $\Delta E_i$  is the width of each channel. The energy interval covered is  $E \in [10.6, 34.9]$  keV. The details of the energy table can be found in the data set at CSA. This choice of energy channels to integrate over is discussed in the Appendix A.

We have also calculated ion number fluxes for the same intervals (not shown in Figure 1). For each energy channel, we calculate the number flux

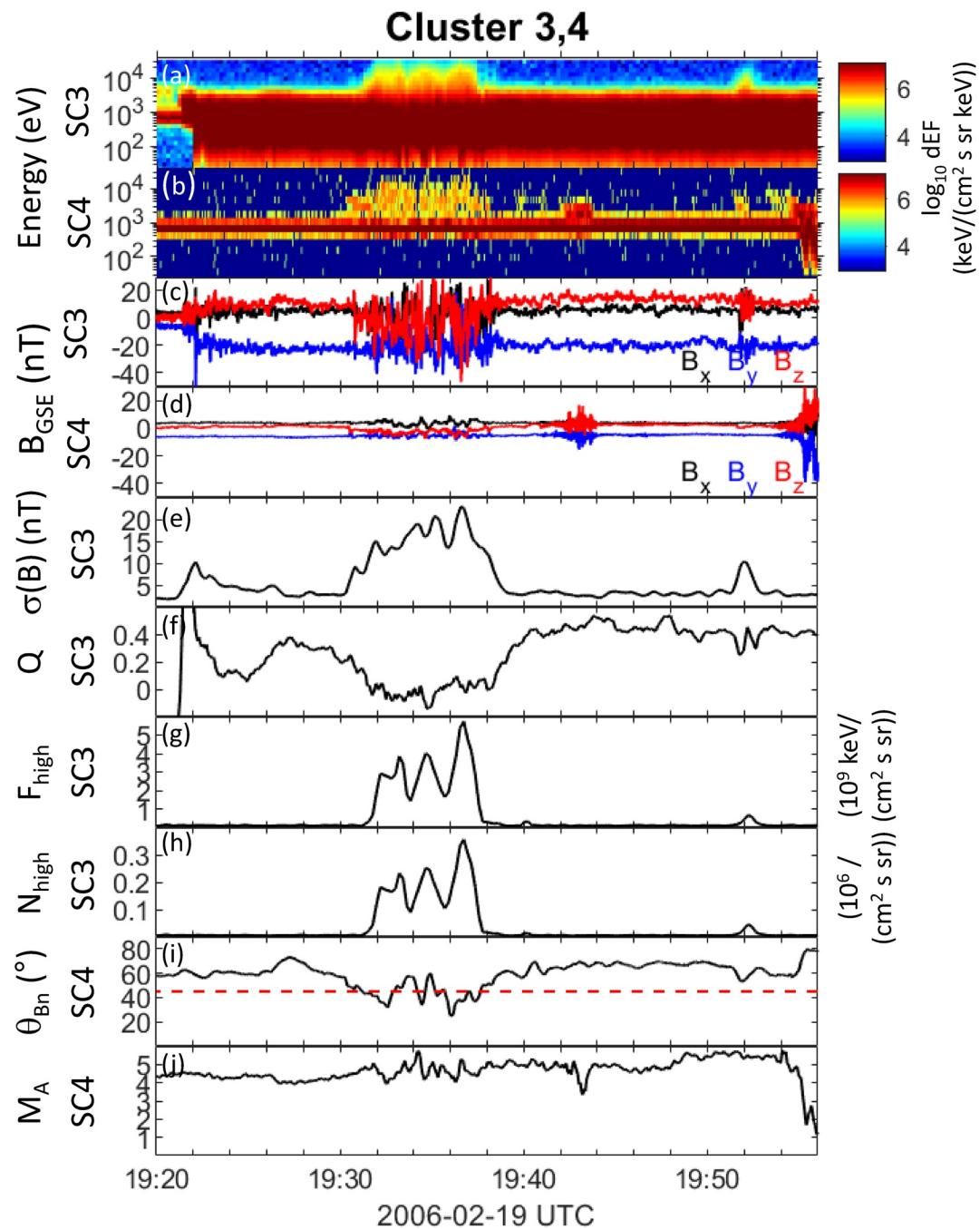
$$N(E_i) = \frac{F(E_i)}{E_i} \Delta E_i \quad (4)$$

and the number flux for high-energy ions

$$N_{high} = \sum_{n=1}^4 N(E_i) \quad (5)$$

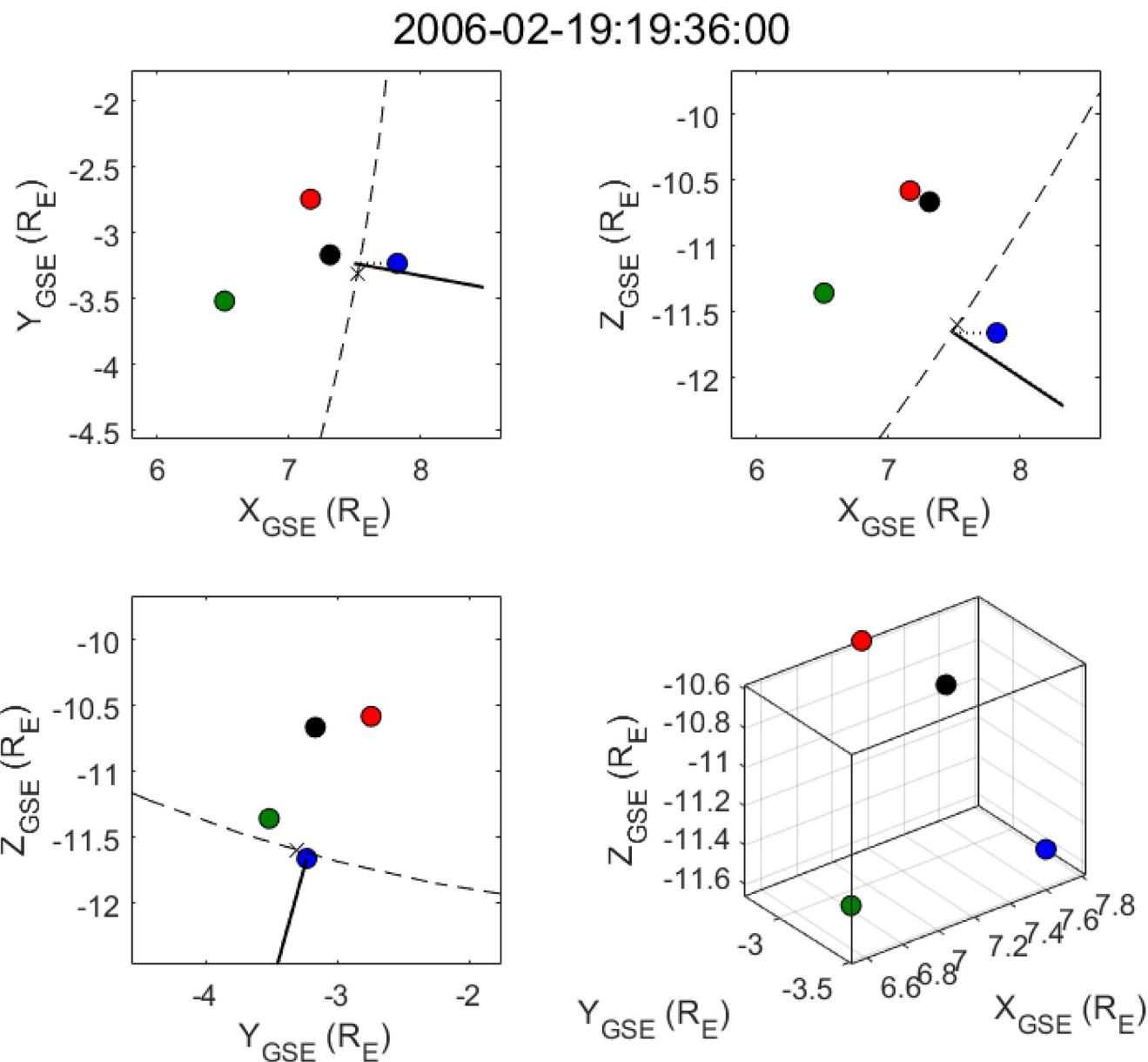
Both  $F_{high}$  and  $N_{high}$ , as well as  $\sigma(\mathbf{B})$  and  $Q$ , are smoothed with a 30-s window.

We have also estimated  $\theta_{Bn}$  from the S/C 4 magnetic field data as follows. We determine the time of the closest bow shock crossing of S/C 4 for the time interval by inspecting the CODIF ion data. For this time interval, we set it to 19:54:30 UTC. We then use the position of S/C 4 for this time to fit a paraboloidal model of the bow shock (Merka et al., 2003). Figure 2 shows the positions of the four Cluster spacecraft at 19:36:00 UTC, approximately in the center of the time interval of interest, in different GSE projections. The resulting fit of the bow shock is shown in the figure. The position of the later intersection of C4 with the bow shock is shown by a cross. We then estimate the position on the bow shock where the solar wind will encounter the shock after first passing the spacecraft, by simply projecting the spacecraft position at 19:36 UTC along the  $x$  axis. At this point, we calculate the normal to the bow shock from the fitted paraboloidal model. Using this normal, we can calculate  $\theta_{Bn}$ , which is then smoothed with a running 30-s window. This type of method has been shown by Horbury et al. (2002) to be more reliable than, e.g., the coplanarity method.



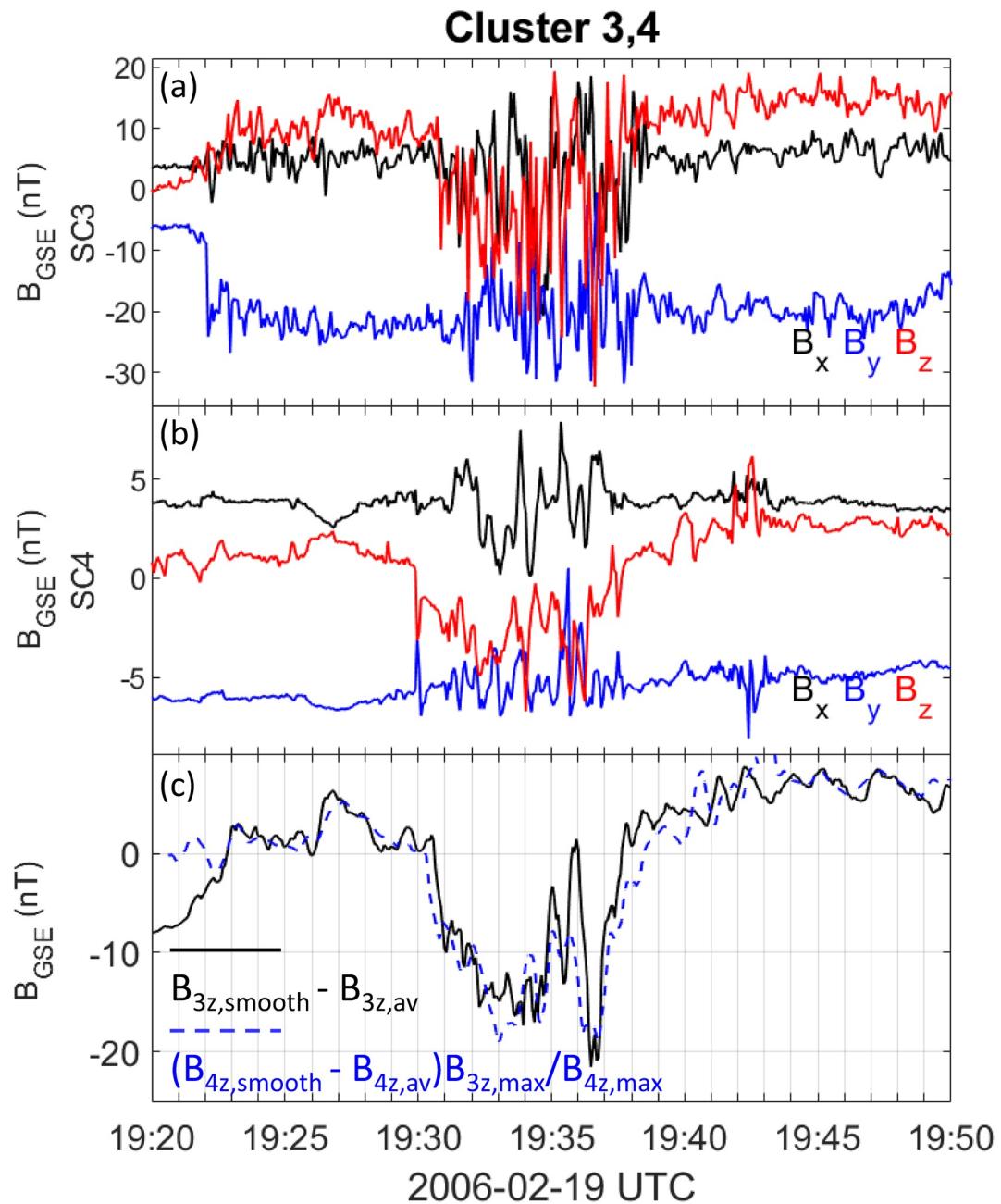
**Figure 1.** Example of Cluster data used in this study. (a) Hot Ion Analyzer ion spectrogram, S/C 3, (b) Composition and Distribution Function ion spectrogram, S/C 4, (c and d) magnetic field in GSE coordinates for S/C 3 and 4, (e) running standard deviation of the magnetic field, S/C 3, (f) ion temperature anisotropy, S/C 3, (g) energy flux of high-energy ions, S/C 3, (h) ion number flux, S/C 3, (i) angle between solar wind magnetic field and bow shock normal, S/C 4, and (j) upstream Alfvén Mach number, S/C 4.

Finally, note that the CODIF ion data for S/C 4 are mainly used for the identification of solar wind and magnetosheath regions, and associated bow shock crossings, with the exception of also calculating the upstream Alfvén Mach number ( $M_A$ ), where we have used CODIF values for ion flow velocity and number density and have assumed that the solar wind contains only protons.  $M_A$  is also smoothed with a 30-s window.



**Figure 2.** The positions of the four Cluster spacecraft in GSE coordinates, for the time 2006-02-19:19:35:00 UTC. The color coding is as follows: S/C 1: black, S/C 2: red, S/C 3: green, and S/C 4: blue. The dashed line shows a paraboloidal model of the bow shock (e.g., Merka et al., 2003), fitted to the bow shock crossing of S/C 4 at 19:54:30 UTC (showed by a small cross in the figure). The dotted line shows a projection along the  $x$  direction from S/C 4, representing where the solar wind will encounter the bow shock after first crossing the spacecraft. At this point, the normal to the bow shock is calculated from the bow shock model and is shown by the solid line, in the different projections.

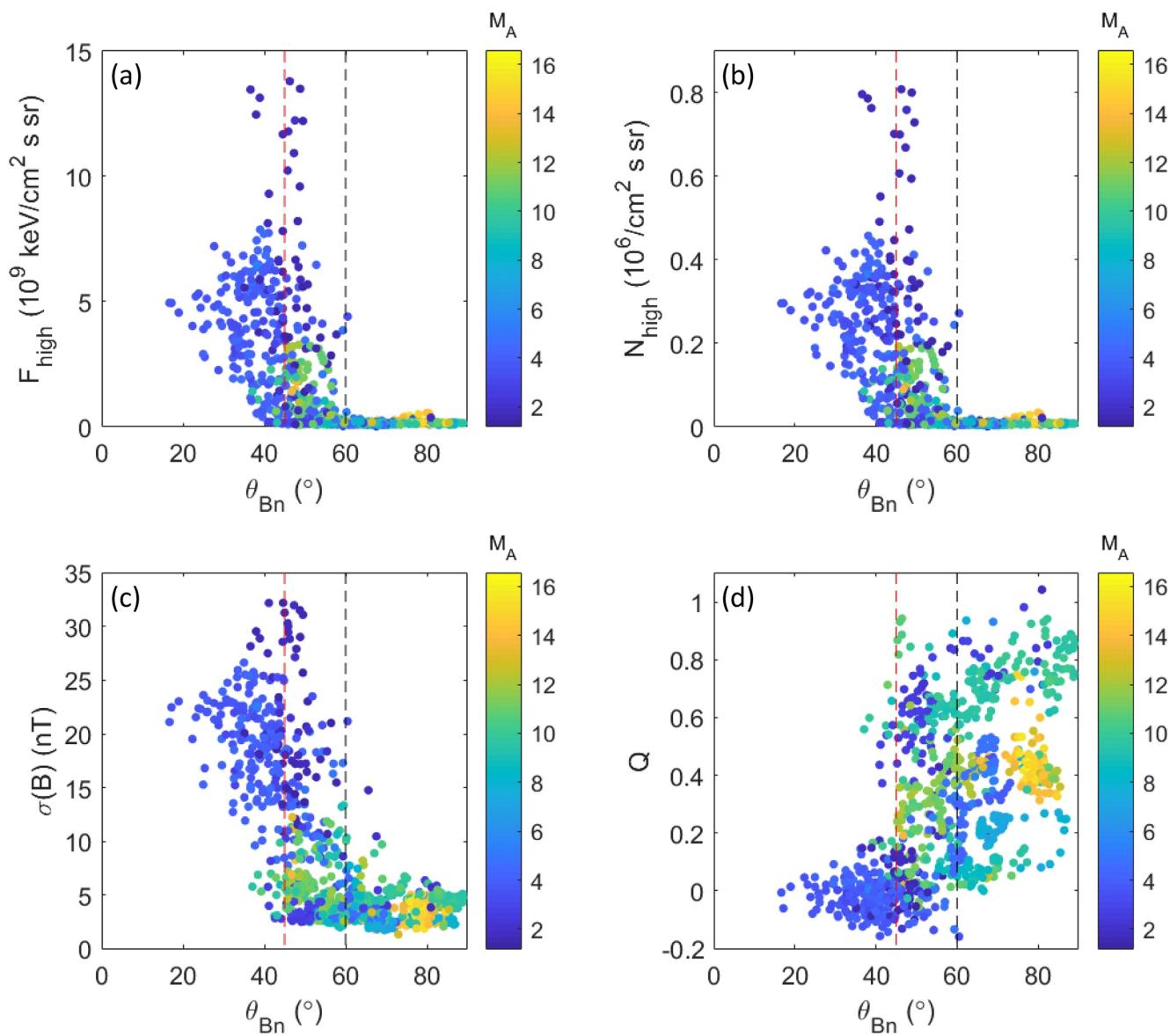
In order to correlate the magnetosheath properties with  $\theta_{Bn}$ , we need to estimate the propagation time for the plasma flow between the solar wind and magnetosheath spacecraft. In order to do this, we manually identify some type of magnetic field structure that can be clearly recognizable in both the solar wind and magnetosheath data, such as a discontinuity or a localized increase or decrease in any of the components or the magnitude of the magnetic field. This usually entails first smoothing the magnetic field to suppress the turbulence and wave activity. A lag is then determined by maximizing the cross correlation between the data from the two different spacecraft. Figure 3 shows the results of this procedure for the time interval shown in Figure 1. The resulting lag time of 36 s is then used to shift the data from C4 in Figure 1, so that a direct comparison between C3 and C4 data can be made.



**Figure 3.** Example of the magnetic field data used to determine the time shift between S/C 3 and 4. (a and b) magnetic field in GSE coordinates for S/C 3 and 4. (c) The smoothed  $z$ -components,  $B_z$  of the magnetic field for S/C 3 (black line) and 4 (broken blue line). For both spacecraft, the average of  $B_z$  has been subtracted. For S/C 4, the result has also been multiplied by the ratio of the maximums of the averaged fields, to facilitate a comparison between the two spacecraft. The resulting magnetic field for S/C 4 has also been shifted by 36 s, which is the lag that maximizes the cross-correlation coefficient between  $B_z$  for the two spacecraft for this time interval.

### 3. Results

From Figure 1, already several interesting observations can be made. Starting from  $\sim$ 19:31 UTC, a region of high-energy ions can be seen in the solar wind, lasting until  $\sim$ 19:38 UTC. This coincides with the time interval where  $\theta_{Bn}$  has a lower value than in the surrounding regions, being close to or below the value of  $45^\circ$ , typical for the quasi-parallel bow shock, although some suprathermal ions can be observed also somewhat earlier, especially in the solar wind.



**Figure 4.** Relationships between the solar wind magnetic field angle to the bow shock normal,  $\theta_{Bn}$ , and various magnetosheath quantities, defined in Section 2. The Mach number,  $M_A$ , of the upstream solar wind for each data point is indicated by the color scale.

From panel (a) in Figure 1, it is clear that the high-energy ions are transmitted into the magnetosheath. This is also clearly seen in the increases in ion energy flux (panels (g and h)). Clearly correlated with the decrease in  $\theta_{Bn}$  is also an increase in  $\sigma(\mathbf{B})$ , and a decrease in the ion temperature anisotropy, consistent with the results of Luhmann et al. (1986) and Dimmock et al. (2015). These correlations can serve as a basis for a local classification of the magnetosheath plasma according to the upstream solar wind conditions.

### 3.1. Relationship With $\theta_{Bn}$

In order to pursue the possibility of such a classification, we show the dependence of the magnetosheath parameters calculated above on  $\theta_{Bn}$  for the whole data set given in Table 1. We have calculated all the quantities for these time intervals in the same way as described in Section 2. All quantities are then resampled at the sample times of the quantities with the lowest time resolution, which are the ion fluxes. This results in a data set of 977 samples. In Figure 4, we have plotted the calculated magnetosheath quantities as a function

of  $\theta_{Bn}$  for all data points. In order to check if the upstream Alfvén Mach number has any influence on the results, we have indicated its value for each data point with color.

It is clear (see panel (a)) that for  $\theta_{Bn} > 60^\circ$ , the high-energy ion energy flux,  $F_{high}$ , is virtually absent. For  $\theta_{Bn} < 45^\circ$ ,  $F_{high}$  is typically greater than the fluxes for large angles, although not for every data point. Interestingly, in the interval of  $45^\circ < \theta_{Bn} < 60^\circ$ —fluxes are sometimes higher than the background level of higher angles, while there is also a population of low fluxes. This is consistent with the fact that the ion foreshock can extend up to  $\theta_{Bn} = 70^\circ$ . The presence of data points with both low and higher energy may be a consequence of the variability of the ion foreshock boundary reported by Le and Russell (1992). It is also possible that the presence or not of high-energy ions depends on the time history of processes in the foreshock, and not only on the instantaneous value of  $\theta_{Bn}$ . Finally, this spread could depend on the properties of the pristine solar wind, such as the Mach number (e.g., Treumann, 2009). While the small values of  $\theta_{Bn}$  are associated with the smaller Mach numbers in panel (a), it is difficult to understand how this could be a real effect. It is likely a consequence of the rather small sample used in this study. For larger  $\theta_{Bn}$  where  $M_A$  covers a larger range, no clear dependence on  $M_A$  can be seen.

The behavior of the number flux  $N_{high}$  (panel (b)) is very similar to that of  $F_{high}$ . From here on, we will only consider  $F_{high}$  and ignore  $N_{high}$ .

Panel (c) shows that the magnetic field standard deviation has a similar behavior as  $F_{high}$  and  $N_{high}$ , with low values for  $\theta_{Bn} > 60^\circ$ , higher values for  $\theta_{Bn} < 45^\circ$ , and a transitional region in between the low and high angles. Finally, the anisotropy  $Q$  (panel (d)), shows a similar but opposite behavior, with low values for low angles, and vice versa. Here, however, two rather clearly defined intervals of the values of  $Q$  for the regions on each side of  $\theta_{Bn} = 45^\circ$  can be seen.

For panels (c and d), of course, low  $\theta_{Bn}$  is also associated with a low Mach number, as in panels (a and b). Again, for larger  $\theta_{Bn}$ , no clear sorting by  $M_A$  can be seen. From this small sample, the dependence on the Mach number is inconclusive and will be ignored in the analysis from now on. We will, however, discuss it briefly again in Section 4.

### 3.2. Classification of Magnetosheath Regions

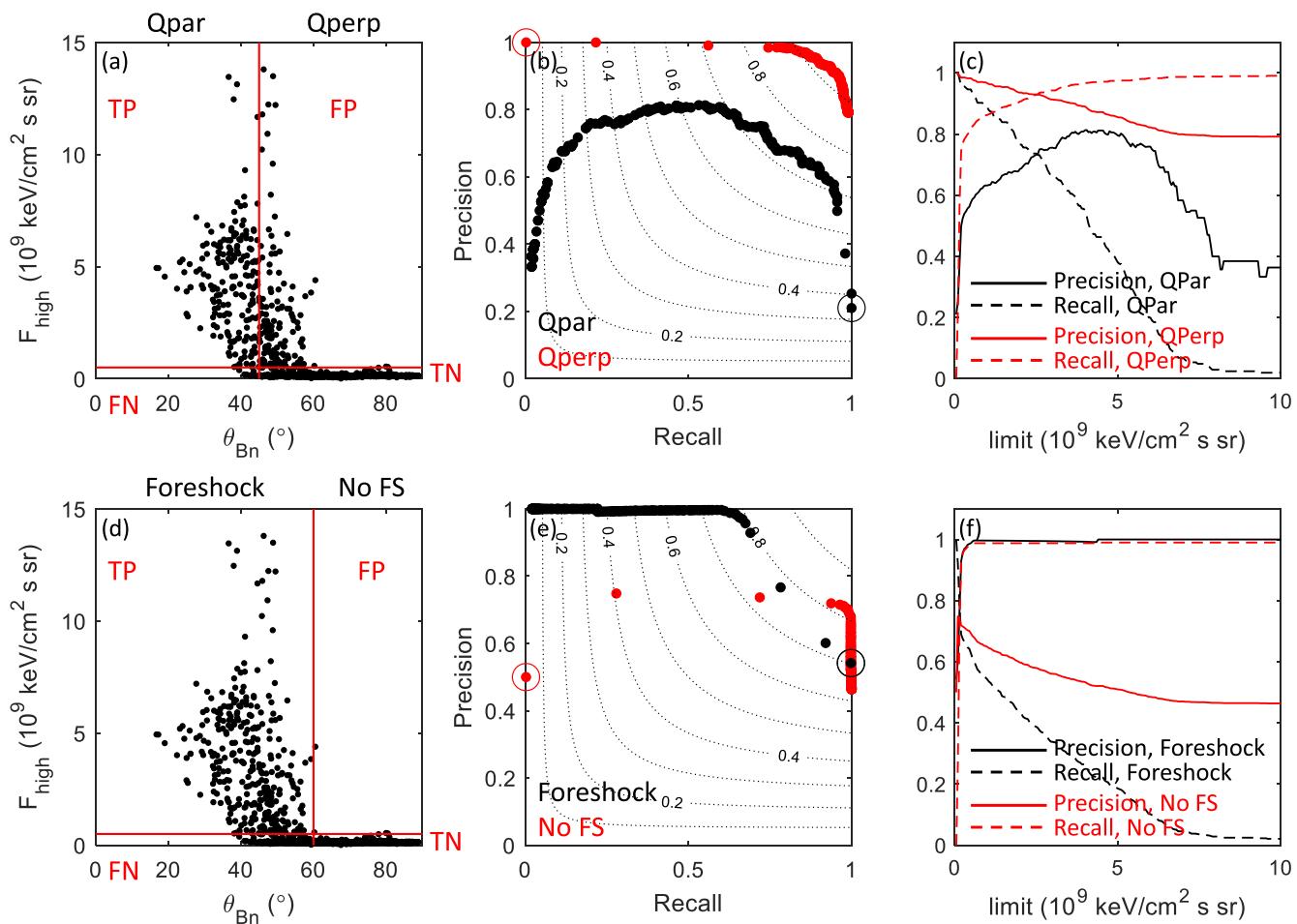
We will now investigate how well the local values of the magnetosheath quantities can be used to classify magnetosheath regions with respect to the upstream conditions. We will here concentrate on a classification of each data point in isolation. The combination of such classifications to whole regions will be touched on very briefly in the Discussion.

Beginning with the high-energy ion energy flux,  $F_{high}$ , let us define two classes, to which each data point can belong "quasi-parallel bow shock" (or "Qpar" for short), meaning that they are associated with a  $\theta_{Bn}$  of less than  $45^\circ$ , and "quasi-perpendicular bow shock" ("Qperp") when  $\theta_{Bn} > 45^\circ$ . This classification based on  $\theta_{Bn}$  is considered to be the "correct" or "true" classification.

We will also then classify every data point according to its value of  $F_{high}$ . For a particular limit of  $F_{high}$ , which we can call  $F_{high}^*$ , we classify a data point as Qpar if  $F_{high} > \text{limit}_{F_{high}}$ , while it is classified as Qperp if  $F_{high} < \text{limit}_{F_{high}}^*$ .

This is illustrated in panel (a) of Figure 5, where the lines  $\text{limit}_{F_{high}} = 0.02 \cdot 10^{-9} \text{ keV}/(\text{cm}^2 \text{ sr})$  and  $\theta_{Bn} > 45^\circ$  divide the plot in four regions. A data point is now referred to as a "true positive" classification of a Qpar magnetosheath data point if  $F_{high} > \text{limit}_{F_{high}}$ , while at the same time  $\theta_{Bn} < 45^\circ$ , i.e., the classifications based on both  $F_{high}$  and  $\theta_{Bn}$  result in the classification of the data point as Qpar. False negatives, false positives, and true negatives are defined analogously in Table 2. The number of true positive data points in the data set is denoted as TP, and for the other cases FP, TN, and FN, respectively. The corresponding regions are indicated in panel (a).

Panel (a) also serves as an illustration of the so-called confusion matrix.



**Figure 5.** Evaluation of magnetosheath classification using the high-energy ion energy flux  $F_{high}$ . (a) The same as Figure 4 but with a limit  $limit_{F_{high}} = 0.02 \cdot 10^9 \text{ keV}/(\text{cm}^2 \text{ s sr})$ , and  $\theta_{Bn} > 45^\circ$  dividing the plot into four regions. (b) Precision–recall plot for the Qpar and Qperp classification, with curves of constant F1 score. (c) Precision and recall for the two classifications as a function of the limit  $limit_{F_{high}}$ . (d–f) The same as (a–c) but for the Foreshock/No foreshock classifications.

$$\begin{pmatrix} TP & FP \\ FN & TN \end{pmatrix} \quad (6)$$

(e.g., Fawcett, 2006; Tharwat, 2020). Visual inspection of panel (a) gives a rough impression of the relative magnitudes of the matrix elements. The smaller the off-diagonal elements are, the better the classification method has worked.

We now define the terms *Precision* and *Recall* (e.g., Fawcett, 2006; Powers, 2007; Tharwat, 2020) as

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

**Table 2**  
*Definitions of Classification Results for Qpar*

Case	Definition	Number of data points
True positive	$(\theta_{Bn} < 45^\circ) \text{ and } (F_{high} > limit_{F_{high}})$	TP
False negative	$(\theta_{Bn} < 45^\circ) \text{ and } (F_{high} < limit_{F_{high}})$	FN
False positive	$(\theta_{Bn} > 45^\circ) \text{ and } (F_{high} > limit_{F_{high}})$	FP
True negative	$(\theta_{Bn} > 45^\circ) \text{ and } (F_{high} < limit_{F_{high}})$	TN

$$\text{Recall} = \frac{TP}{TP + FN}. \quad (8)$$

The precision, loosely speaking, measures how well the classification method picks out only the true positives, while a high recall reflects a more sensitive classification, picking out more true positives but possibly at the cost of including more false positives. A precision of 1 means that there are no false positives, while a recall of 1 means no false negatives.

There is often a trade-off between these two parameters, in choosing the classification method. This can be seen in panel (b), where precision–recall curves for the classification of magnetosheath data points as Qpar are shown with black dots for different values of  $\text{limit}_{F_{high}}$ , which are varied in the interval  $0.005 - 0.3 \cdot 10^{-9}$  keV/(cm<sup>2</sup> sr). The dot with a surrounding circle corresponds to the lowest value of  $\text{limit}_{F_{high}}$ , and the increasing values of  $\text{limit}_{F_{high}}$  are associated with generally higher values of *Precision* and lower values of *Recall*. Panel (b) also shows the curves of constant *F1* score, the harmonic mean of *Precision* and *Recall* (Tharwat, 2020). The *F1* score is often used to optimize a classification algorithm when *Precision* and *Recall* are considered to be approximately equally important, which of course may not be the case in the classifications considered here. Still it is a useful metric to get a first impression of the effects of choosing particular limits. For a more detailed study of how the precision and recall depend on  $\text{limit}_{F_{high}}$ , this dependency is plotted with the black solid and dashed lines in panel (c).

We can see that in order to get a reasonable precision (say 0.8) for the Qpar classification, a rather high  $\text{limit}_{F_{high}}$  has to be chosen. This is due to the rather high number of FP data points for lower values of  $F_{high}$ . This in turn leads to a recall of less than 0.4. This may still be alright, e.g., for a large database, where a large number of data points will be selected even with a strong requirement on the recall. It is, however, not possible to get a precision much greater than 0.8; however, high  $\text{limit}_{F_{high}}$  is set. Visual inspection of panel (a) on the other hand shows that even misclassified events have a  $\theta_{Bn}$  close to 45°.

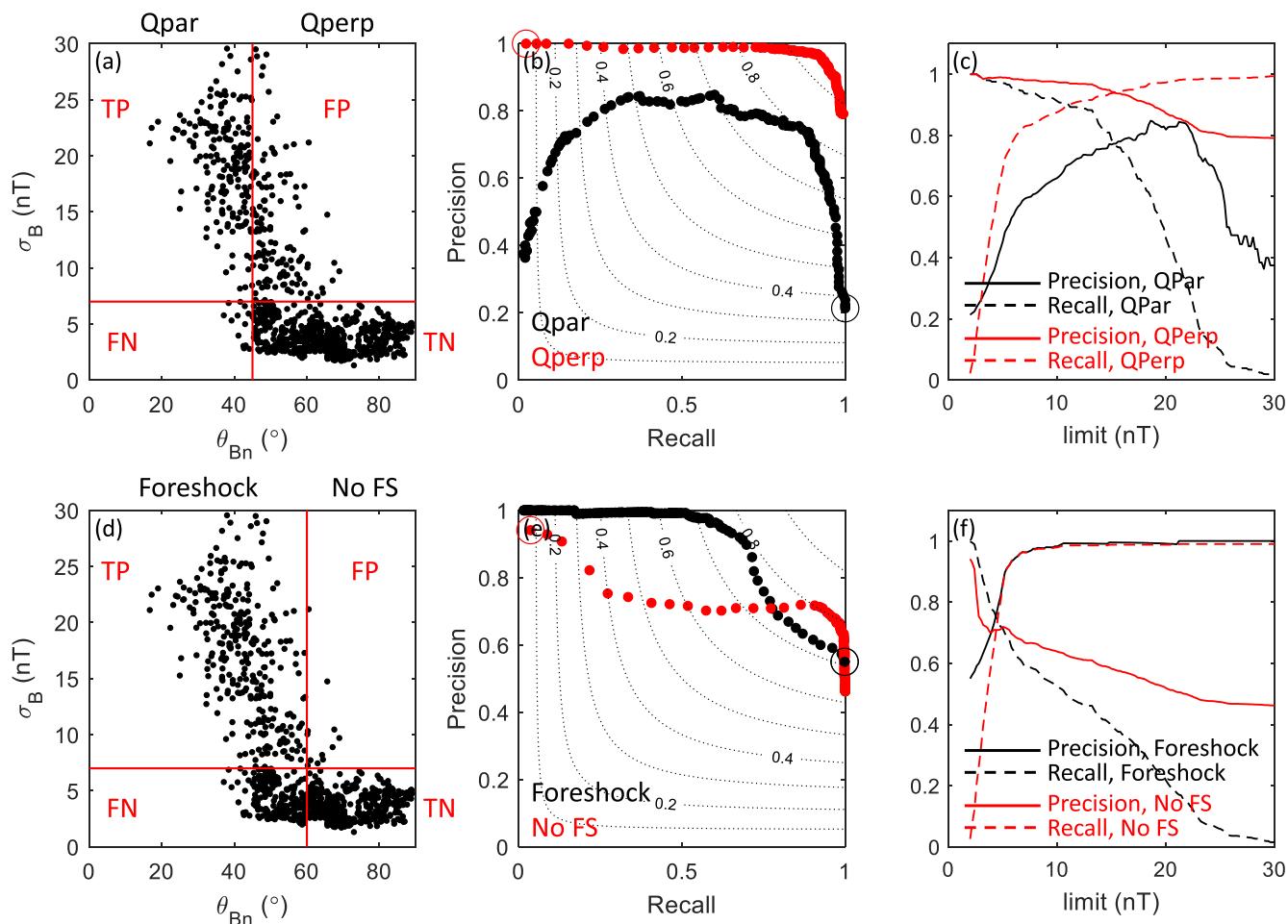
Reversing the meaning of positives and negatives, we can instead study the classification of the magnetosheath data points as Qerp. This is shown in panel (b) as the red dots. Again the point corresponding to the lowest value of  $\text{limit}_{F_{high}}$  is marked by a ring. For the Qerp classification, it is easier to get a high precision, which is consistent with the low number of false positives (in this case, the data points in the lower, left corner of panel (a)). This means that a high value of the recall can be obtained, as also seen in panel (c). We can also note that it is of course not necessary to choose the same value of  $\text{limit}_{F_{high}}$  for the Qpar and Qerp classifications. Choosing different values in this way will result in some data points being unclassified, which may or may not be acceptable.

The fact that high-energy ion fluxes can also be observed for the values of  $\theta_{Bn}$  between 45° and 60° indicates that it can be interesting to classify the data points according to whether the angle is greater than 60° or not. We will classify the data points with  $\theta_{Bn} < 60^\circ$  as Foreshock and the data points with  $\theta_{Bn} > 60^\circ$  as No foreshock. This gives the definitions of TP, etc., similar to Table 2 but for 60° instead of 45°.

Panels (d–f) of Figure 5 were produced analogously to panels (a–c). Here, it can be seen that due to the low number of false positives for the Foreshock class, high precision and recall can easily be obtained. On the other hand, due to the relative large number of data points in the lower left region of panel (d) (false positive with respect to the No foreshock class), the precision remains at around 0.8.

The classification based on the magnetic field standard deviation is shown in Figure 6. The results are rather similar to that of  $F_{high}$ , with a relatively large number of false positives and a lower number of false negatives. This results in similar performances for the precision and recall for the different classifications, except that the precision for Qpar becomes low for the very highest limits.

We finally move on to the classifications based on the ion temperature anisotropy  $Q$  (Figure 7). Due to the opposite sense of the correlation with  $\theta_{Bn}$ , the classification is modified, so that a value of  $Q$  lower than a certain limit  $\text{limit}_Q$  classifies the data point as Qpar. The definitions of true positive, etc., are modified accordingly.

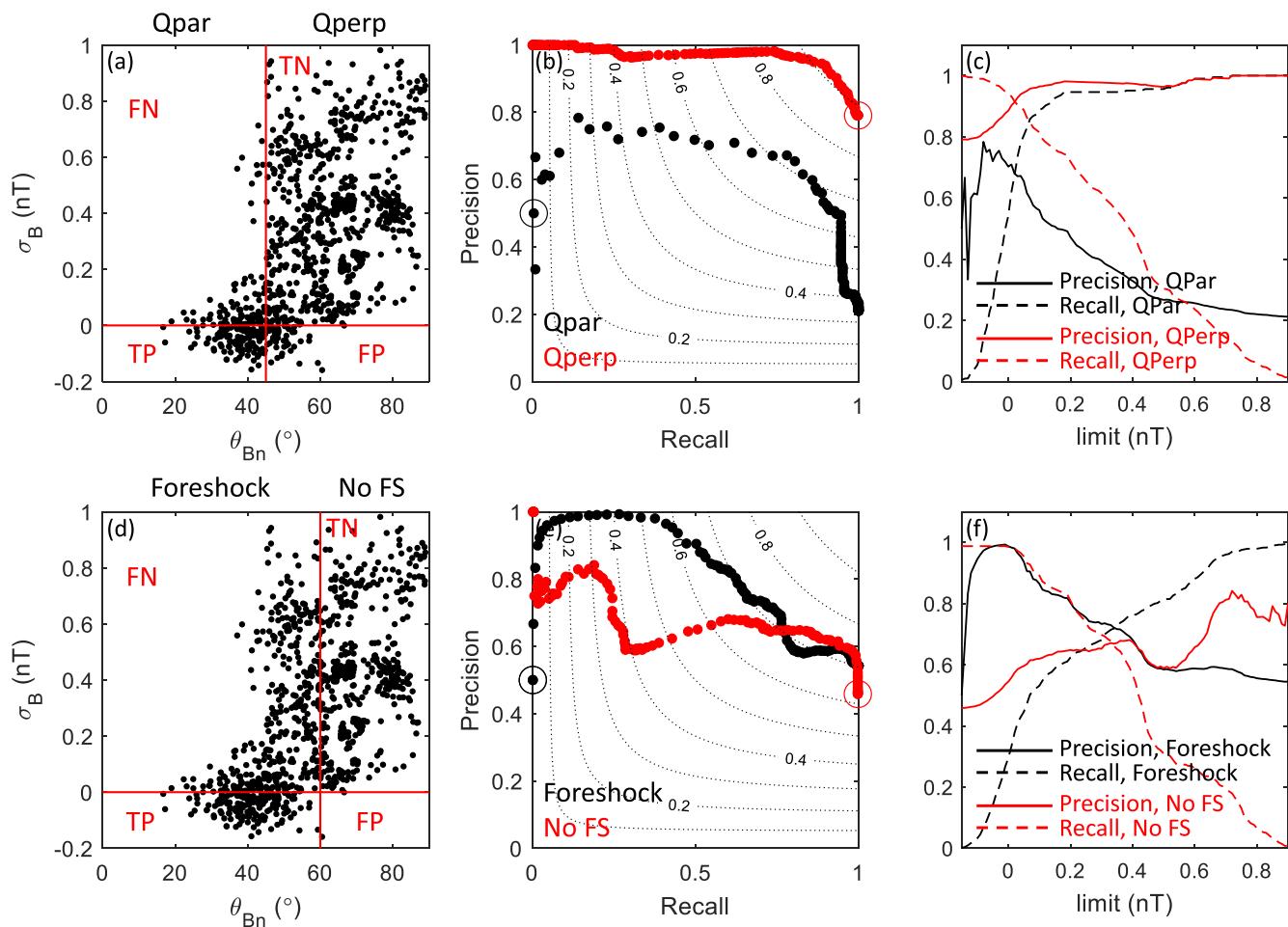


**Figure 6.** Evaluation of magnetosheath classification using the magnetic field standard deviation  $\sigma(B)$  in a format similar to Figure 5.

The classifications of Qpar and Qperp using  $Q$  have in general rather similar properties to those using  $F_{high}$  and  $\sigma(\mathbf{B})$ . A precision of over 0.8 for Qpar can be reached by choosing a low value of  $limit_Q$  (around  $-0.2$ ), at the cost of a rather low recall. For the Qperp classification, high values of both *Precision* and *Recall* can be easily obtained. This is due to the low number of data points in the upper left part of panel (a), which leads to a very low value of false positives for Qperp when  $limit_Q \gtrsim 0.2$ .

Likewise, for the classification of Foreshock, the behavior is similar as for  $F_{high}$  and  $\sigma(\mathbf{B})$ . As long as  $limit_Q$  is set to be greater than around 0, it can be seen in panel (d) that FP becomes zero, leading to *Precision* = 1. On the other hand, however the limit is set, there will still be an appreciable number of false positives for the class No foreshock (the upper-left region), leading to a maximum precision of around 0.8, with a correspondingly low recall.

While classification based on a single parameter can give acceptable results, we may increase the performance by basing the classification on more than one parameter. We show an example of this in Figure 8 for the Qpar classification, noting that the Qperp classification is already very satisfactory. In the figure, we show the results of the classification using both  $F_{high}$  and  $Q$ . The classification is defined in Table 3. The colors of the lines indicated different values of  $limit_{F_{high}}$ , while increasing values of  $limit_Q$  lead to increasing *Recall*, as indicated in the figure. It can be seen that for a rather large range of choices of parameters, values of *Precision* between 0.9 and 1 can be obtained, while still obtaining a *Recall* of over 0.4.



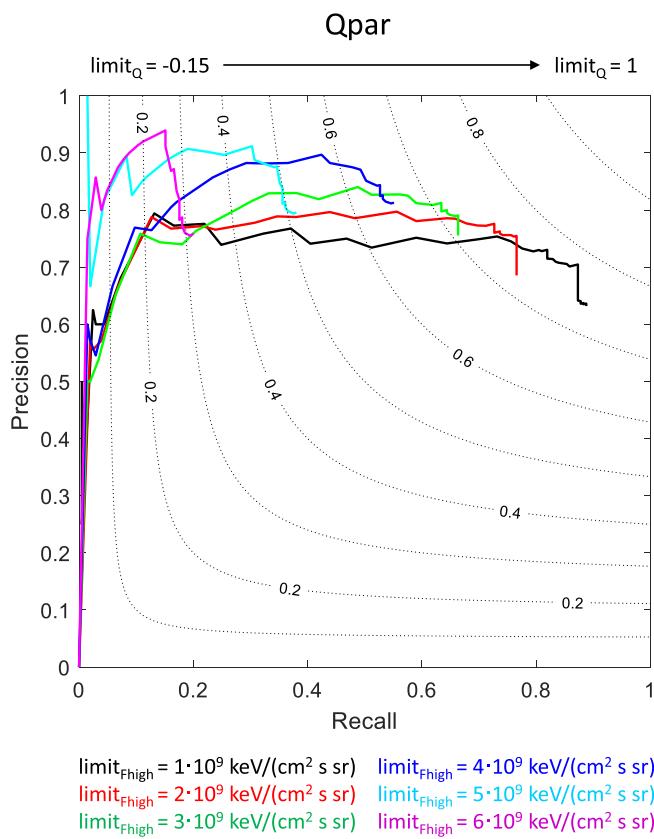
**Figure 7.** Evaluation of magnetosheath classification using the ion temperature anisotropy  $Q$  in a format similar to Figure 5.

### 3.3. Direct Correlations Between High-Energy Ion Fluxes and Magnetic Field Fluctuations in the Magnetosheath and Solar Wind

The results in the above section show that a reliable classification of the upstream conditions based on local magnetosheath measurements can be made with the possible exception of the No foreshock classification. Another approach is, however, possible. While  $\theta_{Bn}$  is often used to classify the upstream conditions, often what is really of interest is the direct manifestations of the foreshock, i.e., the energetic ions and increased magnetic field fluctuations. We will here investigate how well the local magnetosheath high-energy ions reflect the presence of a similar population upstream of the bow shock, and similarly what the relation is between the upstream and downstream magnetic field fluctuations.

To do this, we use the data from S/C 1 and 3, during times where both instruments were operational, and S/C 1 is located in the solar wind, while S/C 3 is in the magnetosheath. Table 4 shows the time intervals that are suitable for such studies. We will here only study two of the time intervals, but list the other times for those interested in pursuing similar studies. Note that here the proximity to the bow shock is not a necessary criterion, since  $\theta_{Bn}$  is not determined.

Figure 9 shows an example of simultaneous solar wind and magnetosheath observations. Panels (a and b) show the ion spectrograms for the two spacecraft in the time interval 2009-03-24:07:40-09:95 UTC. The time series of S/C 1 has been shifted in the same way as those for S/C 4 above. It is clear that the presence of high-energy ions in the solar wind corresponds to a similar population in the magnetosheath. It can also be seen that there is sometimes some additional energization of the ions as they pass the bow shock, e.g., around 07:50 UTC. To take this into account, we can exploit the fact that the temperature of the solar wind



**Figure 8.** Classification of Qpar/Qperp using a combination of  $F_{high}$  and  $Q$ . The color code indicates the  $limit_{Fhigh}$  used. Higher values of  $limit_Q$  correspond to higher Recall.

is much smaller than that of the magnetosheath. The energy of the ions in the pristine solar wind is therefore mainly determined by the ram energy, which is less than  $\sim 5$  keV for protons, even for a solar wind velocity of 1,000 km/s. We therefore extend the integration of the high-energy ion flux in the solar wind to include the six highest energy bins, which extends the integration down to 6.0 keV. The calculation for the magnetosheath is done in the same way as before. The result can be seen in panel (c). There is a good general agreement between the ion energy flux  $F_{high}$  between the upstream and downstream spacecraft, showing again that the local measurements of the high-energy ions are a good indication of the upstream conditions. The correlation is not perfect, which is not surprising, since additional heating, and other processing of the upstream ions when they pass the bow shock is a complicated process that may act differently on different parts of the ion distributions. We can note that there are regions of moderately energized solar wind ions (around 08:15 UTC) that are not transmitted straightforwardly into the magnetosheath.

The correlation of the high-energy fluxes in the solar wind and the magnetosheath can be used for a binary classification in a similar way as in the above sections. We now define TP as the case when the high-energy ion flux in the solar wind,  $F_{high,SW}$ , is greater than a limit  $limit_{SW}$ , and the flux in the magnetosheath,  $F_{high,MSh}$ , is greater than a limit  $limit_{MSh}$ . All the data points are shown in panel (g), together with red lines indicating  $limit_{SW} = 0.4 \cdot 10^9$  keV/(cm<sup>2</sup> s sr) and  $limit_{MSh} = 0.25 \cdot 10^9$  keV/(cm<sup>2</sup> s sr) as an illustrative example. TN, FP, and FN are defined similarly to before, consistent with the regions indicated in panel (g). Here, the off-matrix elements of the confusion matrix are seen to be relatively small, which is promising for further use of this method.

In panels (d–f) are further shown the magnetic field components for the same spacecraft, together with the running standard deviation, defined in the same way as above. Note that the standard deviation for S/C 1 has been multiplied by a factor of 4, in order to facilitate comparison. Panel (h) shows a similar binary classification as for the high-energy ion fluxes, with limits of  $limit_{SW} = 0.5$  nT and  $limit_{MSh} = 5$  nT. It is clear that the correlation between upstream and downstream values is worse for the magnetic field standard deviation than for the high-energy ion fluxes. This may be due to both complexities in how the low-frequency fluctuations are transmitted through the bow shock and the local wave excitation in the magnetosheath.

In Figure 10, we study the binary classification described above for the whole data set of Table 4 in more detail. Panel (a) shows all the data points for the high-energy fluxes in the same format as Figure 9a, but

**Table 3**  
*Definitions of Classification Results for Qpar Using a Combination of  $F_{high}$  and  $Q$*

Case	Definition	Number of data points
True positive	$(\theta_{Bn} < 45^\circ) \text{ and } (F_{high} > limit_{Fhigh}) \text{ and } (Q < limit_Q)$	TP
False negative	$((\theta_{Bn} < 45^\circ) \text{ and } (F_{high} > limit_{Fhigh}) \text{ and } (Q > limit_Q)) \text{ or }$ $((\theta_{Bn} < 45^\circ) \text{ and } (F_{high} < limit_{Fhigh}) \text{ and } (Q < limit_Q)) \text{ or }$ $((\theta_{Bn} < 45^\circ) \text{ and } (F_{high} < limit_{Fhigh}) \text{ and } (Q > limit_Q))$	FN
False positive	$(\theta_{Bn} > 45^\circ) \text{ and } (F_{high} > limit_{Fhigh}) \text{ and } (Q < limit_Q)$	FP
True negative	$((\theta_{Bn} > 45^\circ) \text{ and } (F_{high} > limit_{Fhigh}) \text{ and } (Q > limit_Q)) \text{ or }$ $((\theta_{Bn} > 45^\circ) \text{ and } (F_{high} < limit_{Fhigh}) \text{ and } (Q < limit_Q)) \text{ or }$ $((\theta_{Bn} > 45^\circ) \text{ and } (F_{high} < limit_{Fhigh}) \text{ and } (Q > limit_Q))$	TN

**Table 4**

Data Available to Use for Correlation Between Solar Wind and Magnetosheath High-Energy Ion Energy Flux

Date	Start time	Stop time	SW S/C	MSh S/C	Lag (s)
2006-03-01	07:00	07:12	C1	C3	-7
2009-02-12	00:15	01:40	C1	C3	-8
2009-02-16	16:43	18:05	C1	C3	-17
2009-03-24	07:40	09:05	C1	C3	-120
2009-03-29	03:22	04:16	C1	C3	-109
2009-03-31	11:00	12:30	C1	C3	-75

Note. All times are given in UTC. The spacecraft in the solar wind (SW) and magnetosheath (MSh) are indicated by their standard numbers.

here with  $limit_{SW} = 0.5 \cdot 10^9 \text{ keV}/(\text{cm}^2 \text{ s sr})$  and  $limit_{MSh} = 0.35 \cdot 10^9 \text{ keV}/(\text{cm}^2 \text{ s sr})$ .

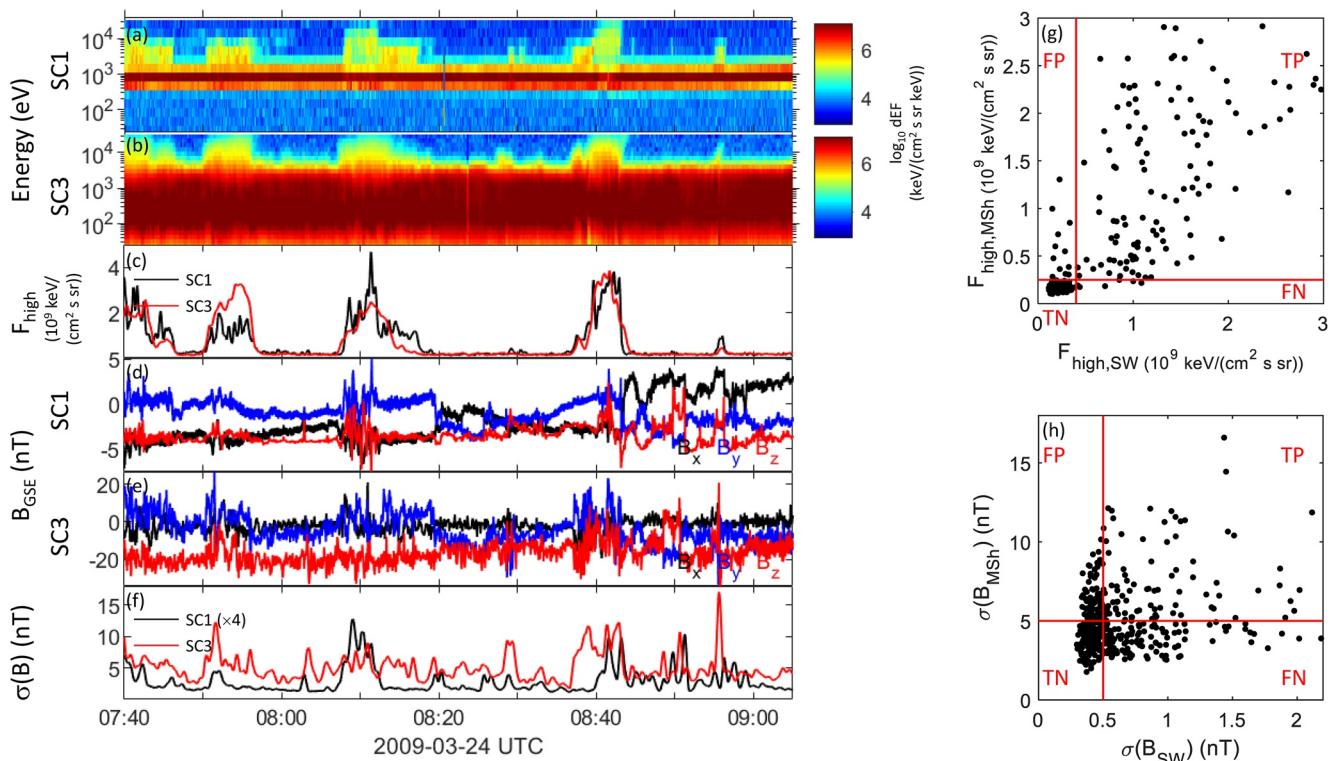
Panel (b) shows a *Precision–Recall* plot, where each colored curve is obtained by varying  $limit_{MSh}$ , while keeping  $limit_{SW}$  constant, with the values indicated by the color. The precision and recall are plotted as a function of  $limit_{MSh}$  in panel (c) for two choices of  $limit_{SW}$ . It can be seen that several choices of limits can result in a combination of quite high values for both the precision and the recall.

Panels (e–g) show similar plots for the magnetic field standard deviation. As suspected from the results in Figure 9, it is more difficult to obtain high values of both the precision and recall here.

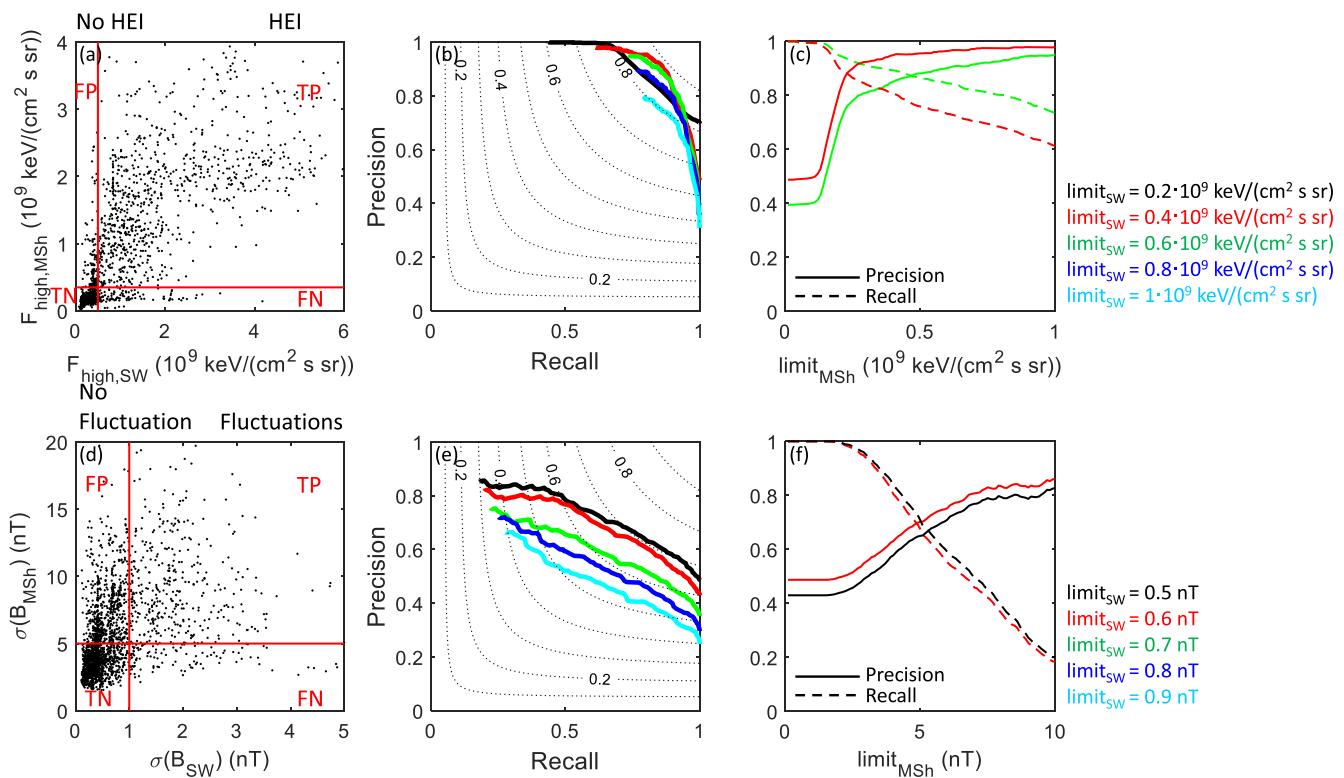
#### 4. Discussion and Future Work

From the investigation above, it seems clear that the local magnetosheath measurements can be used to classify the magnetosheath plasma according to the upstream conditions as measured by  $\theta_{Bn}$ , even on a sample-by-sample basis. We have also shown an example of how the combination of more than one parameter can further increase the quality of the classification. This lends credence to the method used by Raptis, Karlsson, et al. (2020), which did not rely on upstream data. An open question is if their classification corresponds to the Qpar/Qperp or the Foreshock/No foreshock classification examined here, since they did not specify this explicitly. They used a stepwise classification that included several quality rankings, depending on the number of criteria fulfilled. Our interpretation is that their classification corresponds to the aforementioned method of using different limits for the different classification. Taking into account that the parameter range where  $45^\circ < \theta_{Bn} < 60^\circ$  often

ing to the upstream conditions as measured by  $\theta_{Bn}$ , even on a sample-by-sample basis. We have also shown an example of how the combination of more than one parameter can further increase the quality of the classification. This lends credence to the method used by Raptis, Karlsson, et al. (2020), which did not rely on upstream data. An open question is if their classification corresponds to the Qpar/Qperp or the Foreshock/No foreshock classification examined here, since they did not specify this explicitly. They used a stepwise classification that included several quality rankings, depending on the number of criteria fulfilled. Our interpretation is that their classification corresponds to the aforementioned method of using different limits for the different classification. Taking into account that the parameter range where  $45^\circ < \theta_{Bn} < 60^\circ$  often



**Figure 9.** Direct comparisons between solar wind and magnetosheath ion energy fluxes and magnetic fields on 2009-03-24. (a) Ion differential energy flux spectrogram for S/C 1 located in the solar wind. (b) The same for S/C 3 located in the magnetosheath. (c)  $F_{high}$  for S/C 1 and 3. (d and e) Magnetic field for S/C 1 and 3. (f) Magnetic field standard deviation for S/C 1 and 3. (g and h) Binary classification plots for the high-energy flux and magnetic field standard deviation.



**Figure 10.** Evaluation of classifications of the presence of high-energy ions ("HEI") and the high levels of magnetic field fluctuations in a format similar to Figure 5. In panels (b and e), the colors correspond to the different choices of  $limit_{SW}$ , and the increasing values of  $limit_{MSh}$  correspond to the increasing precision.

contains data points, which do not fall into a unique class, their local classification of the magnetosheath plasma likely matches the Qpar classification used here, while their Qperp classification is likely to correspond to the No foreshock classification in this paper. The classification methodology may be a contributing factor to the different relative occurrence of jets behind the quasi-parallel and quasi-perpendicular jets reported by Vuorinen et al. (2019) and Raptis, Karlsson, et al. (2020).

One point of this study is that this difference in classification (Qperp/Qpar vs. Foreshock/No foreshock) can be explicitly discussed. Which version you use may depend on the context. In the study of magnetosheath jets, e.g., you may want to have a clear indication of whether the magnetosheath plasma originates from the bow shock region where large-amplitude bow shock ripples are present. Such ripples are a strong candidate for the generation of those jets (e.g., Hietala & Plaschke, 2013). This would likely correspond to the Qpar class defined here, due to the presence of ULF waves and the nonlinear structures developing from them deeper inside the foreshock. On the other hand, it may be of interest to study the regions connecting to the ion foreshock region with or without upstream ULF waves, in order to investigate if the magnetosheath turbulence has a local or upstream source. This would correspond to magnetosheath plasma classified as Foreshock, but at the same time being either Qpar or Qperp. In view of the results of Section 3.3, an alternative is of course to simply use the level of high-energy ion flux as a direct characterization of the upstream plasma, which is seen to work well for a rather broad range of limits. However, the direct correlation between the upstream and downstream magnetic field fluctuations is likely to be disturbed by fluctuations generated locally in the magnetosheath, as mentioned above.

While a reasonable precision can be obtained for the individual data points, in practice, it may be more desirable to classify the larger regions of the magnetosheath. In order to do so, one could use some type of clustering/postprocessing algorithm. A simple example could be to use a window of a certain temporal length and to assign the class according to the majority classification of the data points for the whole time interval within the window. Such a method could be applied in several steps with increasing window widths, until a desired granularity has been obtained. Many other methods are possible, of course.

While the results here are encouraging for the goal of using local magnetosheath measurements to characterize the upstream conditions, several subjects still remain to be investigated.

In this study, we have used Cluster configurations with satellite separations of around 1–2  $R_E$ . In one respect, this is a strength of the study, since it unambiguously relates the upstream and downstream data to each other. However, it is still unclear if this methodology can be applied to the magnetosheath plasma further downstream, in the flanks, or if mixing and thermalization of the plasma, and the local excitation of waves will affect the reliability of the method. An indication that it will still work comes from inspection of the plasma in this region, which still exhibits clearly defined regions with and without high-energy ions associated with different magnetic field orientations and fluctuation levels and different temperature anisotropies. To study this further, we will correlate measurements between Cluster spacecraft with similar separations to those in this study, but with both e.g., S/C 1 and S/C 3 located in the magnetosheath. We can then determine if the correlation progressively changes or is still similar to the ones observed here. An alternative method is to correlate Cluster measurements with, e.g., THEMIS (Angelopoulos, 2009) or MMS measurements, for times when one of the spacecraft is in the near-Earth solar wind (and example can be seen in Karlsson et al., 2018), and another in the magnetosheath, in order to increase the satellite separation. This may entail tracing of the magnetosheath plasma to its point of origin at the bow shock, which could be done using analytical models of magnetosheath flow (e.g., Kallio & Koskinen, 2000). A third possibility would be to use only THEMIS spacecraft to correlate upstream and downstream properties, when one spacecraft is in the solar wind and another in the magnetosheath.

While the dependence on the upstream Mach number was inconclusive in this study, it is important to study the dependence on other upstream parameters than  $\theta_{Bn}$ . Both the upstreams  $M_A$  and  $\beta$  could affect the results. When those parameters are small, perhaps the energetic ion flux may be below the limit of  $F_{high}$  chosen here, resulting in false negatives. This could be particularly important further downstream in the flank magnetosheath, where the shock is weaker since the normal component of the solar wind velocity is smaller.

Deeper inside the magnetosheath, closer to the magnetopause, the presence of leaked energetic magnetospheric ions (e.g., Sibeck et al., 1987) may also be a complication. Here, it will also be important to consider the other parameters, such as temperature anisotropy and magnetic field variance. We can also speculate that the presence of oxygen ions could be used to discriminate such time periods.

In this study, we have argued that a fixed energy interval to integrate the high-energy flux works well. However, when applying it to times of low solar wind Mach numbers, or at the flanks, the foreshock ions may mostly have energies of less than 10 keV. It may then be necessary to have a more flexible integration range, which could perhaps be determined by the magnetosheath temperature.

When applying the method to other missions, it will be necessary to adjust  $limit_{F_{high}}$ . A useful guide to how to set this value is to argue that it should approximately be the energy flux of the instrumental noise, integrated over the interval chosen, even if locally energized ions in the magnetosheath may also provide false positives. An indirect indication that this method should work reasonably well is that for the illustrative levels chosen in panel 9(g), which were chosen by visual inspection to make the confusion matrix elements small,  $limit_{SW}$  is higher than  $limit_{MSH}$ . Since the integration interval was larger for the solar wind data, this is expected according to the above reasoning.

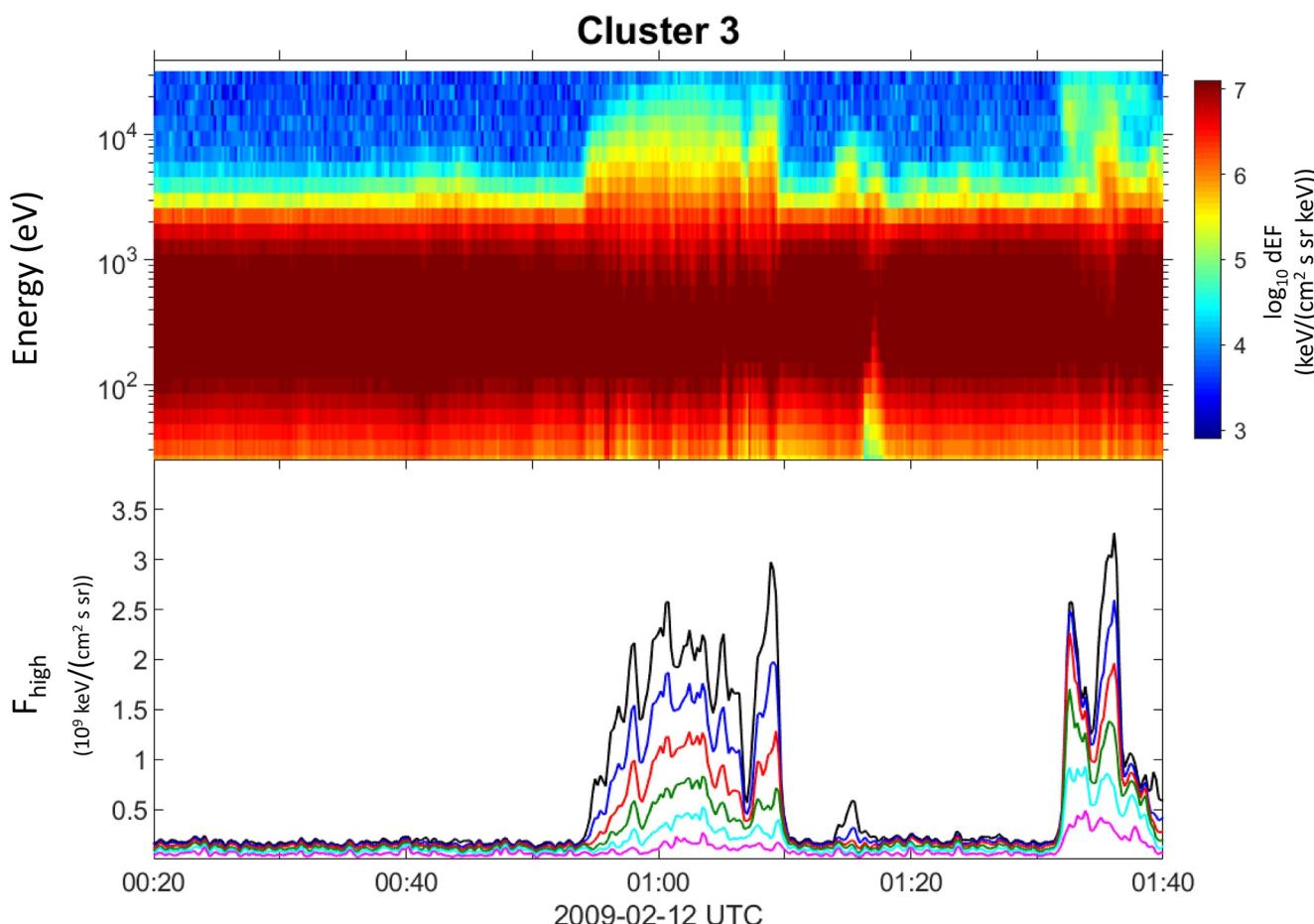
We have only used the omnidirectional ion spectra and have further reduced the information by integrating over an energy interval. There is more information available in the full ion distributions, and as mentioned in the Introduction, different regions of the ion foreshock contain different types of ion populations. Comparing the full ion distributions both in the upstream and magnetosheath plasma and investigating how they are transformed when they cross the bow shock should yield further information of the source of magnetosheath plasma. In particular, the detailed ion distributions could possibly be used to trace their origin with respect to the distance from the ion foreshock boundary.

The classification in this study has been defined in a semimanual way. This is a good methodology for beginning to understand the basic properties of these types of classifications, but it is likely that a more optimal result can be obtained by machine-learning methods, which eliminate human biases and preconceptions (although at times at the expense of transparency). The analysis presented here provides the necessary foundation for implementing machine learning techniques that have been recently used with great success in heliophysics (Camporeale et al., 2018) and to similar classification problems, e.g., the MMS plasma

environment classification (Breuillard et al., 2020; Olshevsky et al., 2019) and magnetosheath jet classification (Raptis, Karlsson, et al., 2020). A direct implementation of the discussed suggestions (distribution functions and multispacecraft conjunctions) to a supervised learning model (e.g., logistic regression and neural networks) could provide improved results by using all the available input in an optimal way. Another approach would be to use unsupervised learning techniques (e.g., Self Organizing Maps (SOMs), clustering algorithms, etc.) that have also been used for classification problems regarding the pristine solar wind (e.g., Amaya et al., 2020; Li et al., 2020) and require no manual preprocessing to take place. We believe that such approaches will be the next step toward characterizing the upstream origin of the plasma found in the magnetosheath and magnetosphere regions, and plan to progress along these lines.

### Appendix A: Setting the Limit in Calculation of $F_{high}$

We here discuss the question of which energy bins to include in the calculation of  $F_{high}$ . This is particularly important during the time period when the magnetosheath temperature is high, and the thermal ions can have energies up to around 10 keV. Including bins with these energies will then result in false positives when identifying the high-energy ions of upstream origin. Therefore, only energies down to ~10 keV should be included in the  $F_{high}$  calculation. In order to check that this is enough, we here show two examples of the high energy flux calculated for various lower limits of the integration. The first example (Figure A1) is from 2012-02-19 and is used in Section 3.3.



**Figure A1.** Ion spectrogram and high-energy flux calculated using different number of bins. The lowest curve (magenta) uses energy bins 1–2, the next one (cyan) uses energy bins 1–3, etc.

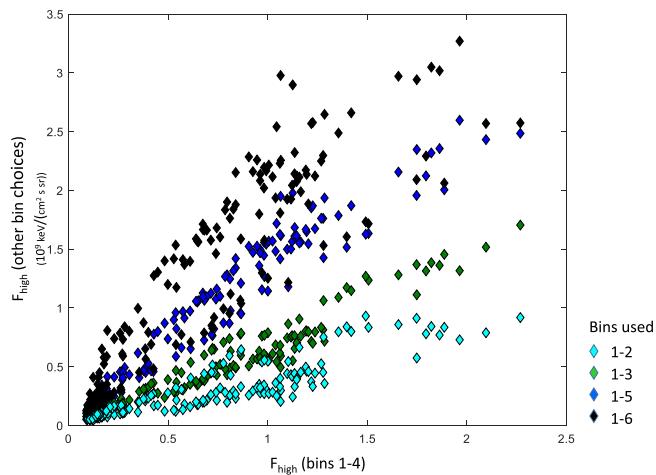
**Table A1***Energy Table for the Highest Six Energy Bins for the CIS HIA Instrument, for the Time Intervals Used in This Study*

Bin no.	Low energy limit (keV)	High energy limit (keV)
1	25.0	34.9
2	18.0	25.0
3	14.1	18.0
4	10.6	14.1
5	8.0	10.6
6	6.0	8.0

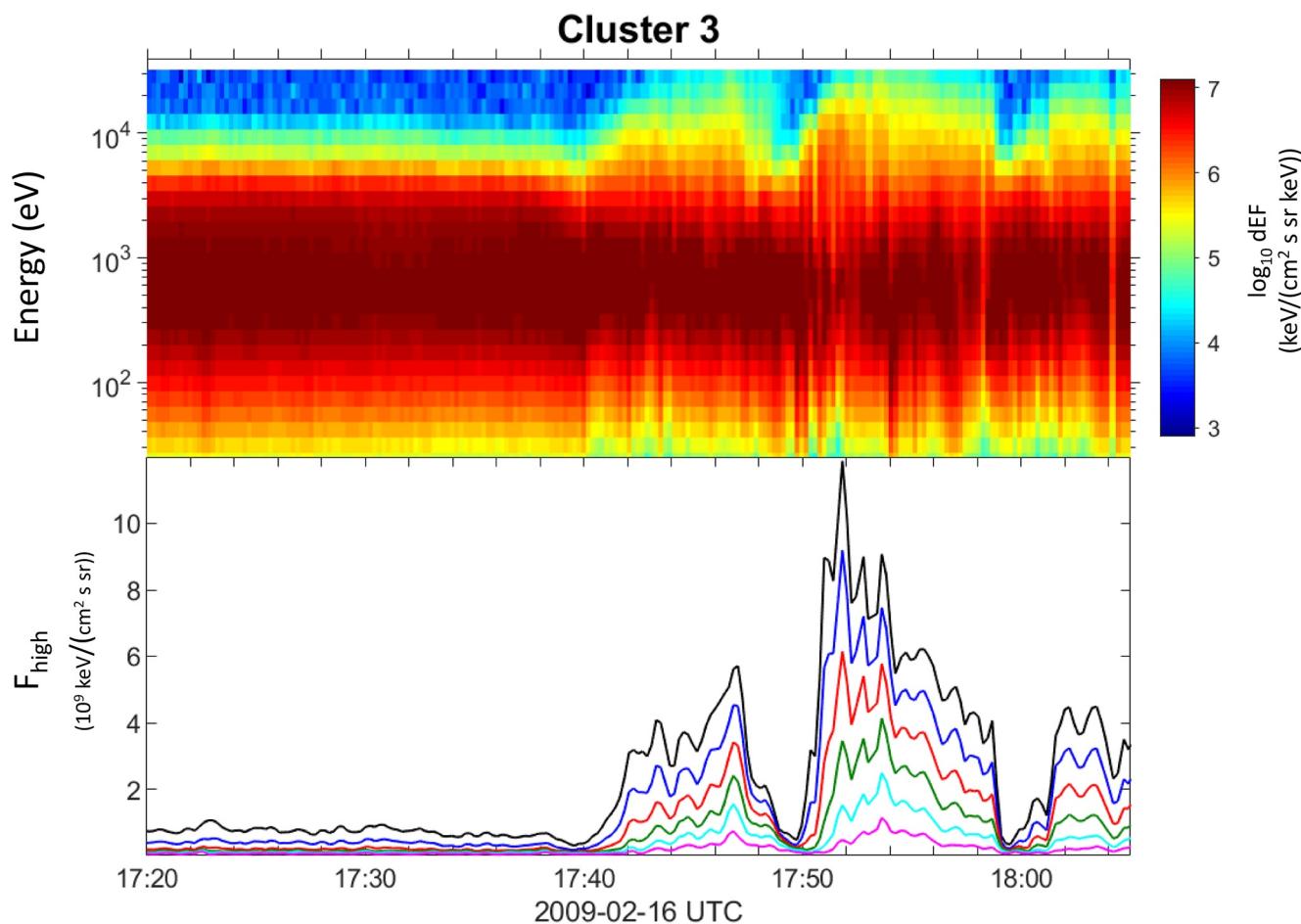
In the second panel of Figure A1 is shown the integrated energy flux using only the highest energy bin (magenta curve with the lowest values), together with the resulting fluxes using the highest 2–6 energy bins, with increasingly higher total fluxes. (The values of the energy limits of the bins are given in Table A1.) As can be seen, all alternatives give similar qualitative results, with a clear separation from the background levels from times where the ion spectrum shows no appreciable high-energy ions (although the separation is marginal when only using the one highest energy bin).

To further check this, we have plotted (Figure A2), for each time, the values of  $F_{high}$  using the different choices of bins to integrate over, vs.  $F_{high}$  calculated for bins 1–4. As seen, there is a strong correlation between the values for different alternatives of the integration interval, meaning that, for this case, with a relatively low magnetosheath temperature, several different choices are possible.

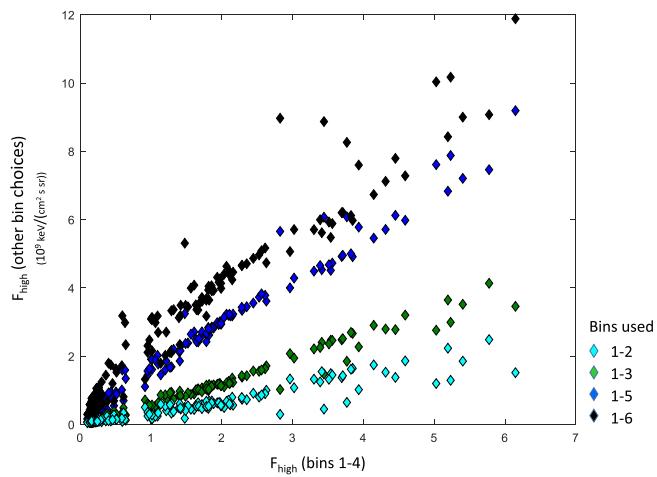
Next, we show similar plots (Figures A3 and A4) for a time interval from 2009-02-16, also used in Section 3.3. Here, the magnetosheath temperature is higher, and the ion energies reach around 10 keV. We can see that this somewhat affects the level of the background values of  $F_{high}$  for the calculations using the energy bins down to below 10 keV (black and blue lines), although the effect is not dramatic. A similar scatter plot as above again shows that there is a good correlation between the results using different integration boundaries, although the effect of a higher background level for the extended integration regions (bin choices 1.5 and 1–6) is visible as a different slope in the curves for  $F_{high} < 1$ .



**Figure A2.** Scatterplot of the results shown in Figure A1. As reference shown on the abscissa, the values are used from the calculation using energy bins 1–4, which are then plotted against the results from calculations using other energy bin intervals.



**Figure A3.** Same format as Figure A1 for a time interval from 2009-02-16.



**Figure A4.** Same format as Figure A2 for a time interval from 2009-02-16.

In conclusion, the choice of the interval of integration is not critical for our purposes in this paper, but the choice of integrating the four highest energy bins seems the logical choice, since the results are robust and are not affected by the thermal magnetosheath ions. This corresponds to including ion energies down to 10.6 keV in the calculation of  $F_{high}$ . For example, THEMIS, this typically leaves three high-energy bins for the integration of  $F_{high}$ , which according to the above analysis likely is fine. As an alternative, one may have to exclude the events with very high magnetosheath temperatures.

## Data Availability Statement

The data in this study are available via the Cluster Science Archive (Laakso et al., 2010) (<https://www.cosmos.esa.int/web/csa>).

## Acknowledgments

SR and TK are supported by the Swedish National Space Agency (SNSA grant 90/17). HT and TK are supported by the Swedish National Space Agency (SNSA grant 190/19).

## References

- Amaya, J., Dupuis, R., Innocenti, M. E., & Lapenta, G. (2020). Visualizing and interpreting unsupervised solar wind classifications. *Frontiers in Astronomy and Space Sciences*, 7, 66. <https://doi.org/10.3389/fspas.2020.553207>
- Angelopoulos, V. (2009). The themis mission. *The THEMIS mission*, 5–34. [https://doi.org/10.1007/978-0-387-89820-9\\_2](https://doi.org/10.1007/978-0-387-89820-9_2)
- Archer, M. O., & Horbury, T. S. (2013). Magnetosheath dynamic pressure enhancements: Occurrence and typical properties. *Ann. Geophys.*, 31, 319–331. <https://doi.org/10.5194/angeo-31-319-2013>
- Balogh, A., Carr, C. M., Acuna, M., Dunlop, M., Beek, T., Brown, P., et al. (2001). The Cluster magnetic field investigation: Overview of in-flight performance and initial results. *Annales Geophysicae*, 19(10/12), 1207–1217. <https://doi.org/10.5194/angeo-19-1207-2001>
- Balogh, A., Schwartz, S., Bale, S., Balikhin, M., Burgess, D., Horbury, T., et al. (2005). Cluster at the bow shock: Introduction. *Space Science Reviews*, 118(1–4), 155–160. <https://doi.org/10.1007/s11214-005-3826-1>
- Battarbee, M., Blanco-Cano, X., Turc, L., Kajdič, P., Johlander, A., Tarvus, V., et al. (2020). Helium in the earth's foreshock: A global vlasitator survey. In *Annales Geophysicae* (Vol. 38, pp. 1081–1099). <https://doi.org/10.5194/angeo-38-1081-2020>
- Breuillard, H., Dupuis, R., Retino, A., Le Contel, O., Amaya, J., & Lapenta, G. (2020). Automatic classification of plasma regions in near-earth space with supervised machine learning: Application to magnetospheric multi scale 2016–2019 observation. *Frontiers in Astronomy and Space Sciences*, 7, 55. <https://doi.org/10.3389/fspas.2020.00055>
- Burch, J., Moore, T., Torbert, R., & Giles, B. (2016). Magnetospheric multiscale overview and science objectives. *Space Science Reviews*, 199(1–4), 5–21. [https://doi.org/10.1007/978-94-024-0861-4\\_2](https://doi.org/10.1007/978-94-024-0861-4_2)
- Camporeale, E., Wing, S., & Johnson, J. (2018). *Machine learning techniques for space weather*. Elsevier.
- Carr, C., Brown, P., Alconcel, L., Oddy, T., Fox, P., & Whiteside, B. (2013). *User guide to the FGM measurements in the cluster active archive (CAA)*. Retrieved from <http://caa.estec.esa.int>
- Chang, S.-W., Scudder, J., Fennell, J., Friedel, R., Lepping, R., Russell, C., et al. (2000). Energetic magnetosheath ions connected to the earth's bow shock: Possible source of cusp energetic ions. *Journal of Geophysical Research*, 105(A3), 5471–5488. <https://doi.org/10.1029/1999ja900468>
- Crooker, N., Eastman, T., Frank, L., Smith, E., & Russell, C. (1981). Energetic magnetosheath ions and the interplanetary magnetic field orientation. *Journal of Geophysical Research*, 86(A6), 4455–4460. <https://doi.org/10.1029/ja086ia06p04455>
- Dandouras, I., & Barthe, A. (2012). *User guide to the CIS measurements in the Cluster Active Archive (CAA)*. CIS Team.
- Dimmock, A., Osmane, A., Pulkkinen, T., & Nykyri, K. (2015). A statistical study of the dawn-dusk asymmetry of ion temperature anisotropy and mirror mode occurrence in the terrestrial dayside magnetosheath using themis data. *Journal of Geophysical Research: Space Physics*, 120, 5489–5503. <https://doi.org/10.1002/2015ja021192>
- Eastwood, J., Lucek, E., Mazeille, C., Meziane, K., Narita, Y., Pickett, J., & Treumann, R. (2005). The foreshock. *Space Science Reviews*, 118(1–4), 41–94. <https://doi.org/10.1007/s11214-005-3824-3>
- ESA. (2020). *ESA Science and Technology - mission timeline*. Retrieved from <https://sci.esa.int/web/cluster/-/23160-constellation-geometry-over-time>
- Escoubet, C., Schmidt, R., & Goldstein, M. (1997). *The Cluster and Phoenix missions* (pp. 11–32). Springer. [https://doi.org/10.1007/978-94-011-5666-0\\_1](https://doi.org/10.1007/978-94-011-5666-0_1)
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fuselier, S., Klumpar, D., & Shelley, E. (1991). On the origins of energetic ions in the earth's dayside magnetosheath. *Journal of Geophysical Research*, 96(A1), 47–56. <https://doi.org/10.1029/90ja01751>
- Fuselier, S., Lewis, W., Schiff, C., Ergun, R., Burch, J., Petrinec, S., & Trattner, K. (2016). Magnetospheric multiscale science mission profile and operations. *Space Science Reviews*, 199(1–4), 77–103. [https://doi.org/10.1007/978-94-024-0861-4\\_4](https://doi.org/10.1007/978-94-024-0861-4_4)
- Fuselier, S. A. (1994). Suprathermal ions upstream and downstream from the earth's bow shock. *Solar Wind Sources of Magnetospheric Ultra-Low-Frequency Waves*, 81, 107–119.
- Gary, S. P. (1991). Electromagnetic ion/ion instabilities and their consequences in space plasmas: A review. *Space Science Reviews*, 56(3–4), 373–415. <https://doi.org/10.1007/bf00196632>
- Goncharov, O., Gunell, H., Hamrin, M., & Chong, S. (2020). Evolution of high-speed jets and plasmoids downstream of the quasi-perpendicular bow shock. *Journal of Geophysical Research: Space Physics*, 125, e2019JA027667. <https://doi.org/10.1029/2019JA027667>
- Hietala, H., & Plaschke, F. (2013). On the generation of magnetosheath high-speed jets by bow shock ripples. *Journal of Geophysical Research: Space Physics*, 118, 7237–7245. <https://doi.org/10.1002/2013JA019172>
- Horbury, T., Cargill, P., Lucek, E., Eastwood, J., Balogh, A., Dunlop, M., et al. (2002). Four spacecraft measurements of the quasiperpendicular terrestrial bow shock: Orientation and motion. *Journal of Geophysical Research*, 107(A8). SSH-10. <https://doi.org/10.1029/2001ja000273>
- Kallio, E. J., & Koskinen, H. E. (2000). A semiempirical magnetosheath model to analyze the solar wind-magnetosphere interaction. *Journal of Geophysical Research*, 105(A12), 27469–27479. <https://doi.org/10.1029/2000ja000086>

- Karlsson, T., Plaschke, F., Hietala, H., Archer, M., Blanco-Cano, X., Kajdič, P., et al. (2018). Investigating the anatomy of magnetosheath jets–MMS observations. In *Annales Geophysicae* (Vol. 36, pp. 655–677). <https://doi.org/10.5194/angeo-36-655-2018>
- Kis, A., Scholer, M., Klecker, B., Kucharek, H., Lucek, E., & Rème, H. (2007). Scattering of field-aligned beam ions upstream of earth's bow shock. In *Annales Geophysicae* (Vol. 25, pp. 785–799). <https://doi.org/10.5194/angeo-25-785-2007>
- Laakso, H., Perry, C., McCaffrey, S., Herment, D., Allen, A., Harvey, C., et al. (2010). *The Cluster Active Archive* (pp. 3–37). Springer. [https://doi.org/10.1007/978-90-481-3499-1\\_1](https://doi.org/10.1007/978-90-481-3499-1_1)
- Le, G., & Russell, C. (1992). A study of ulf wave foreshock morphology—I: Ulf foreshock boundary. *Planetary and Space Science*, 40(9), 1203–1213. [https://doi.org/10.1016/0032-0633\(92\)90077-2](https://doi.org/10.1016/0032-0633(92)90077-2)
- Li, H., Wang, C., Tu, C., & Xu, F. (2020). Machine learning approach for solar wind categorization. *Earth and Space Science*, 7(5), e2019EA000997. <https://doi.org/10.1029/2019ea000997>
- Luhmann, J., Russell, C., & Elphic, R. (1986). Spatial distributions of magnetic field fluctuations in the dayside magnetosheath. *Journal of Geophysical Research*, 91(A2), 1711–1715. <https://doi.org/10.1029/ja091ia02p01711>
- Merka, J., Szabo, A., Narock, T., King, J., Paularena, K., & Richardson, J. (2003). A comparison of IMP 8 observed bow shock positions with model predictions. *Journal of Geophysical Research*, 108(A2). <https://doi.org/10.1029/2002ja009384>
- Olshevsky, V., Khotyaintsev, Y. V., Divin, A., Delzanno, G. L., Anderzen, S., Herman, P., et al. (2019). *Automated classification of plasma regions using 3d particle energy distribution*. arXiv preprint arXiv:1908.05715.
- Plaschke, F., Hietala, H., & Angelopoulos, V. (2013). Anti-sunward high-speed jets in the subsolar magnetosheath. *Annales Geophysicae*, 31, 1877–1889. <https://doi.org/10.5194/angeo-31-1877-2013>
- Plaschke, F., Karlsson, T., Hietala, H., Archer, M., Vörös, Z., Nakamura, R., et al. (2017). Magnetosheath high-speed jets: Internal structure and interaction with ambient plasma. *Journal of Geophysical Research: Space Physics*, 122, 10157–10175. <https://doi.org/10.1002/2017JA024471>
- Powers, D. M. (2007). *Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation*.
- Raptis, S., Aminalragia-Giamini, S., Karlsson, T., & Lindberg, M. (2020). Classification of magnetosheath jets using neural networks and high resolution omni (hro) data. *Frontiers in Astronomy and Space Sciences*, 7. <https://doi.org/10.3389/fspas.2020.00024>
- Raptis, S., Karlsson, T., Plaschke, F., Kullen, A., & Lindqvist, P.-A. (2020). Classifying magnetosheath jets using mms: Statistical properties. *Journal of Geophysical Research: Space Physics*, 125, e2019JA027754. <https://doi.org/10.1029/2019ja027754>
- Réme, H., Bosqued, J., Sauvaud, J., Cros, A., Dandouras, J., Aoustin, C., et al. (1997). The Cluster ion spectrometry (CIS) experiment. In *The Cluster and Phoenix missions* (pp. 303–350). Springer. [https://doi.org/10.1007/978-94-011-5666-0\\_12](https://doi.org/10.1007/978-94-011-5666-0_12)
- Sibeck, D., McEntire, R., Lui, A., Lopez, R., Krimigis, S., Decker, R., et al. (1987). Energetic magnetospheric ions at the dayside magnetopause: Leakage or merging? *Journal of Geophysical Research*, 92(A11), 12097–12114. <https://doi.org/10.1029/ja092ia11p12097>
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.
- Treumann, R. (2009). Fundamentals of collisionless shocks for astrophysical application, 1. Non-relativistic shocks. *Astronomy and Astrophysics Review*, 17(4), 409–535. <https://doi.org/10.1007/s00159-009-0024-2>
- Tsurutani, B. T., & Rodriguez, P. (1981). Upstream waves and particles: An overview of isee results. *Journal of Geophysical Research*, 86(A6), 4317–4324. <https://doi.org/10.1029/ja086ia06p04317>
- Vuorinen, L., Hietala, H., & Plaschke, F. (2019). Jets in the magnetosheath: Imf control of where they occur. In *Annales Geophysicae* (Vol. 37, pp. 689–697). <https://doi.org/10.5194/angeo-37-689-2019>
- Wilson, L. (2016). *Low Frequency Waves at and Upstream of Collisionless Shocks*. Washington DC American Geophysical Union Geophysical Monograph Series (Vol. 216, pp. 269–291). <https://doi.org/10.1002/9781119055006.ch16>