JGR Space Physics

RESEARCH ARTICLE
10.1029/2024JA033323

Can XMM-Newton Be Used to Track Compositional Changes in the Solar Wind?

S. Nitti[1], J. A. Carter[1], S. F. Sembay[1], S. E. Milan[1], L. Zhao[2], S. T. Lepri[2], and K. D. Kuntz[3]

[1]School of Physics and Astronomy, University of Leicester, Leicester, UK, [2]Department of Atmospheric, Oceanic, and Space Sciences, University of Michigan, Ann Arbor, MI, USA, [3]The Henry A. Rowland Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, USA

Key Points:

- Variable solar wind charge exchange is associated with periods of outlier wind having increased solar wind proton flux
- XMM-Newton observes differences in solar wind charge exchange emission count rate during different solar wind types
- The distribution of the OVIII line flux from XMM-Newton is similar to that of the ACE O^{8+}/p ratio across different solar wind types

Correspondence to:
S. Nitti,
sn364@leicester.ac.uk

Citation:
Nitti, S., Carter, J. A., Sembay, S. F., Milan, S. E., Zhao, L., Lepri, S. T., & Kuntz, K. D. (2024). Can XMM-Newton be used to track compositional changes in the solar wind? *Journal of Geophysical Research: Space Physics*, *129*, e2024JA033323. https://doi.org/10.1029/2024JA033323

Received 16 SEP 2024
Accepted 2 DEC 2024

Abstract Geocoronal Solar Wind Charge Exchange (SWCX) is the process by which heavy ions from the solar wind undergo charge exchange with neutral hydrogen atoms from the Earth's exosphere, releasing photons at discrete energies characteristic of the solar wind ions. This paper investigates the solar wind types driving geocoronal SWCX. We find that during periods of time-variable SWCX, higher fractions of every ion species are recorded by ACE compared to the averages. Notably, a subset of the slow solar wind characterized by a systematic lower temperature and higher proton flux is surprisingly effective for producing SWCX. Given the degradation of the solar wind composition spectrometer on ACE in 2011, we explore the capabilities of XMM-Newton as an alternative sensor to monitor heavy ion composition in the solar wind. Unlike the distributions of other ion line fluxes analyzed, only OVIII, extracted via spectral analysis of XMM-Newton observations, display patterns similar to the corresponding parent ion abundances from ACE $\left(O^{8+}/p\right)$. Finally, we employ a Random Forest Classifier model to predict solar wind types based on literature results. When combining proton data with XMM-Newton features, the model performance improves significantly, achieving a macro-averaged F1 score of 0.80 (with a standard deviation of 0.06).

1. Introduction

Heavy ions of the solar wind refer to any element heavier than He, that can be completely, or almost completely, stripped of electrons. They consist primarily of C, N, O, Ne, Ni, Mg, Si, S, and Fe, and constitute approximately 1% of the overall solar wind population (e.g., Gloeckler & Geiss, 1989). Their abundances are determined low in the solar corona (Bame et al., 1974), and reflect the local electron temperature and density history of the wind ions before they freeze-in (Hundhausen, 1972). From the freeze-in point, the solar plasma travels outwards through interplanetary space keeping its composition mostly unaltered (Hundhausen et al., 1968).

While traveling through the Solar System, the solar wind encounters obstacles such as the Earth's magnetic field, which acts as an obstruction to its supersonic flow. This interaction causes the solar wind to abruptly decelerate, resulting in the formation of a shock front known as the bow shock. Downstream of the bow shock, the dynamic pressure of the solar wind transforms into thermal pressure, that is, the temperature, pressure, and density of the solar wind increase. The shocked solar wind plasma cannot mix with Earth-based plasma, so it flows around the magnetopause, forming a region called the magnetosheath. The magnetosheath plasma is often turbulent and exhibits small-scale variations in density, magnetic field strength, and velocity (Rakhmanova et al., 2016, and references therein). The high solar wind density and the presence of neutral hydrogen from the exosphere, the outermost layer of the Earth's atmosphere, causes this region to be a source of soft X-rays ($\lesssim$2 keV) due to solar wind charge exchange (SWCX; Cravens, 1997; Freyberg, 1998). SWCX is the process in which highly ionized solar wind heavy atoms collide with neutral atoms and acquire an electron in an excited state. The decay of the electron to a lower energy state causes the emission of a photon in the UV or soft X-ray bands. The process is formally described by the equations:

$$X^{q+} + M \rightarrow X^{*(q-1)+} + M^{+}, \tag{1}$$

$$X^{*(q-1)+} \rightarrow X^{(q-1)+} + h\nu, \tag{2}$$

where X^{q+} is the parent heavy ion carried by the solar wind, and M is the neutral atom. The asterisk on top of $X^{(q-1)+}$ denotes the excitation state of the newly acquired electron. To clarify conventions between emission lines and ions, OVIII emission lines result from the de-excitation of O^{*7+}, which is itself generated through charge exchange from the solar wind *parent* ion, O^{8+}. This explains the focus in this paper on comparisons between parent ions and their corresponding charge exchange emission lines.

SWCX occurs not only in the Earth's exosphere, but it is ubiquitous in the Solar System, arising wherever the solar wind interacts with interstellar neutrals flowing through the heliosphere (e.g., Lallement, 2004), or planetary environments like Mars, Venus, Jupiter, and Pluto (see Branduardi-Raymont, 2022; Bhardwaj, 2006, and references therein). Accurate modeling of the SWCX emission is a crucial aspect of understanding the interactions between the solar wind heavy ions and the local neutral environment. Generally, the SWCX emissivity at a point **x** is approximated by Equation 3 (Cravens, 1997),

$$Q(\mathbf{x}) = \alpha\, n_H\, n_{SW}\, v_{SW}, \tag{3}$$

$$\alpha_H(V) = \sum_{X^{q+}{}_j} \sigma_{X^{q+},H}(V) b_{X^{(q-1)+}{}_j} E_j \frac{X^{q+}}{p} \tag{4}$$

where, n_H is the neutral hydrogen density, $n_{SW}\, v_{SW}$ is the solar wind heavy ion flux at a point **x**, and α is the efficiency factor of the charge exchange. The α parameter (Equation 4) is proportional to the abundances of the solar wind heavy ions involved in the process, $\frac{X^{q+}}{p}$, their cross-section with neutral hydrogen at their collision velocity, $\sigma_{q+,H}(V)$, the branching ratio, $b_{X^{(q-1)+}{}_j}$, which is the fraction of ions with charge state $(q-1)+$ that relax through transition j, and the energies for each emitted transition line, E_j. The heavy ion measurements in Equation 4 are often taken from the Advanced Composition Explorer (ACE; Stone et al., 1998), located at the Lagrange point L1, then shifted in time to the Earth's bow shock (e.g., Carter et al., 2011; Koutroumpa, 2024; Whittaker et al., 2016). This workaround is due to the current lack of heavy ion measurements in the vicinity of Earth.

One way to validate SWCX models is by comparing simulations to observations of SWCX emissions made by X-ray telescopes, for example, by XMM-Newton, Suzaku, Chandra, and Swift. Among them, only XMM-Newton (Jansen et al., 2001) allows narrow field-of-view observations of SWCX emission from the magnetosheath. However, XMM-Newton is an astronomical telescope and is not optimized for viewing SWCX. It does so serendipitously, only when its line of sight traverses the magnetosheath while pointing at astronomical targets (more in Section 2). Although comparisons between XMM-Newton observations and SWCX model simulations show a generally positive correlation, there can be extreme variations from case to case (e.g., Carter et al., 2011; Whittaker et al., 2016). Reasons for such variations might be found in a possible variability of the exospheric density as a result of a strong geomagnetic activity (Cucho-Padin & Waldrop, 2019), or fluctuations in solar wind composition within the turbulent magnetosheath that current models may not fully capture. It should also be noted that, since the field of view of XMM is so small (15'), a small uncertainty in the location of the magnetopause can make a very large difference in whether strong SWCX emission is seen or not. Given that MHD models do not generally agree on the location of the magnetopause to better than $\pm 1\ R_E$, strong disagreements between model and observation can occur even when the model is essentially correct.

In this paper, we investigate the viability of utilizing heavy ion ACE data shifted from L1 to the bow shock as an approximation for the turbulent magnetosheath composition when modeling SWCX emissions. This is done by comparing the heavy ion composition data from ACE with the heavy ion line fluxes derived from a previous archival spectral analysis of XMM-Newton observations that showed evidence of time-variable geocoronal SWCX. Our goals include.

1. Investigate the predominant solar wind types and properties responsible for driving geocoronal SWCX.
2. Assess the extent to which the solar wind variability observed by ACE is reflected in XMM-Newton data.

By doing this, we will explore *the feasibility of using XMM-Newton as an alternative solar wind heavy ion monitor within the inner magnetosphere*, particularly given the degradation of the solar wind ion composition spectrometer (SWICS instrument) onboard ACE in 2011, due to a radiation and age-induced hardware anomaly

(see SWICS 2.0 Level 3 Version 1.11 Data Release Notes (https://izw1.caltech.edu/ACE/ASC/DATA/level2/ss2/swics_lv2_V2_release_notes.txt)). Since then, no further heavy ion measurements have been made from L1 or closer to Earth. Such measurements are crucial for improving our understanding of the SWCX process and supporting future X-ray missions, including the upcoming SMILE mission (Branduardi-Raymont et al., 2018). SMILE, a joint mission by the European Space Agency and the Chinese Academy of Sciences, is scheduled for launch in 2025 and will capture large-scale magnetopause dynamics for the first time using soft X-ray imaging (Sembay et al., 2024).

This paper is organized as follows: in Section 2, we describe the data sets used for our analysis. In Section 3, we measure the plasma properties of the solar wind during 1998–2011 and compare it with times when XMM detected SWCX emission from the Earth's exosphere (henceforth SWCX periods) between 2000 and 2009. In Section 4, we analyze to what extent we can use XMM-Newton data to classify the solar wind types driving SWCX in absence of ion measurements from ACE. In Section 5, we discuss our results, summarize the conclusions and further work.

2. Data Set

2.1. XMM-Newton List of Exospheric SWCX

XMM-Newton, hereafter XMM, is a spacecraft launched in 1999 into a highly elliptical orbit, with an apogee of about 15 R_E (Earth Radii) and a perigee of ca. 4 R_E, which takes it far below the dayside ecliptic plane. As it points to its more distant astrophysical targets, XMM occasionally observes through the dayside magnetosheath, where most of the SWCX emission can be found. The XMM X-ray imaging suite of cameras is known as the European Photon Imaging Camera (EPIC), and is comprised of two Metal Oxide Semi-conductor (MOS; Turner et al., 2001) and one pn (Strüder et al., 2001) charged coupled device (CCD) arrays. Both cameras complement each other in terms of characteristics and capabilities, allowing for comprehensive X-ray observations over an energy range from 0.2 to 15 keV (even though calibration at energies below ∼0.35 keV is highly questionable). This covers the SWCX band (below ∼2 keV) and higher energies for accurate monitoring of the non-SWCX background. It is crucial to note that XMM does not provide in situ measurements of X-ray emission. Instead, it integrates the observed emission along its line-of-sight (LOS), based on the selected energy range.

For our analysis, we use a set of 103 XMM/EPIC observations exhibiting time-variable exospheric SWCX emissions, found in data between the years 2000 and 2009 (Carter et al., 2011). The total coverage of the data set is approximately 1,000 hr, with each observation lasting from 2 hr to 1 day. These SWCX events were found through temporal variability in the diffuse low-energy X-ray emission, ranging from 0.5 to 0.7 keV, compared to a non-variable continuum between 2.5 and 5.0 keV, the *background*. The low-energy band is mainly dominated by OVIII and OVII emission lines, which are key indicators of charge exchange emission, and the continuum was assumed to be free of exospheric SWCX. All of the chosen lines-of-sight are known to pass through or near the nose of the magnetosheath, thus the significant temporal variability of this set could only be explained by exospheric emission, ruling out other heliospheric sources. Heliospheric contributions from interstellar neutrals fluctuate over much longer timescales and exhibit less intensity variation (Robertson & Cravens, 2003). They provide a nearly constant offset to the background, which Carter et al. (2011) removed from the data set via spectral background subtraction.

The Carter et al. (2011) data set comprises low-energy lightcurves and the average ion line fluxes produced via charge exchange of the solar wind parent ions: O^{7+}, O^{8+}, C^{5+}, C^{6+}, and Mg^{11+}. The ion line fluxes were extracted through spectral fitting analysis, employing a free fitting approach at fixed energies, determined by the theoretical model of Bodewits et al. (2007) for charge exchange interactions between H-like and He-like ions and hydrogen atoms. The fitting process accounted for contributions from both background and instrumental spectra.

2.2. ACE Data of the Solar Wind at L1

ACE is a spacecraft located at the Lagrange point L1 between the Sun and Earth, which provides in situ measurements of the incoming solar wind and interplanetary magnetic field (IMF). From the Solar Wind Ion Composition Spectrometer (SWICS; Gloeckler et al., 1992), we employ the ACE/SWICS 1.1 2-hr Level 2 data set, which covers the period between 1998 and 2011. In particular, we take the elemental abundances relative to oxygen $\frac{X}{O}$, the ion charge state distributions $\frac{X^{q+}}{X}$, and the He^{++} number density. From the ACE/SWICS 12-min

Level 2 data set, we take the number density of solar wind protons (p). We also use ACE/MAG (Magnetic Field Experiment; Smith et al., 1998) measurements of the total magnitude of the IMF, and ACE/SWEPAM (Solar Wind Electron, Proton and Alpha Monitor; McComas et al., 1998) 1-Hour Level 2 data for the solar wind velocity and the spacecraft position components. We shifted all the data from L1 to Earth, using the time-shifting equation from the OMNI 2 data set on the OMNIWeb page (https://omniweb.gsfc.nasa.gov/html/ow_data.html). This approach provides a good approximation of the data time-shift with an accuracy of a few minutes, which is suitable given the time resolution of the ACE data set.

3. Method

We compare the solar wind heavy ion evolution from L1 to Earth, by answering the following questions.

1. Which solar wind types are the main drivers of geocoronal SWCX?
2. Does the distribution of solar wind properties during time-variable SWCX reflect those of the entire ACE data set from 1998 to 2011?
3. Does the distribution of the heavy ion composition as measured by ACE/SWICS reflect the distribution of the ion fluxes extracted from XMM-Newton data during SWCX periods?
4. Is the variability observed in the SWCX emission from XMM-Newton associated with a particular solar wind type?

3.1. Solar Wind Classification

We use the solar wind classification methods for ACE data gathered by Koutroumpa (2024) from literature results. It splits the solar wind into five categories: "Streamer" slow solar wind, "Outlier" slow solar wind, "Upper Depleted Wind" (UDW), fast solar wind from coronal holes (CHs), and interplanetary coronal mass ejections (ICMEs), (see below for definitions). ICME events are taken from the list of Richardson and Cane (2010), whereas the other solar wind types are based upon the in situ measurement of ion abundances from the ACE/SWICS 1.1 database, which spans from 1998 to 2011.

Figure 1a shows a 2D histogram (adapted from Koutroumpa, 2024, Figure 3) of the occurrence of the ratios $\frac{O^{7+}}{O^{6+}}$ and $\frac{C^{6+}}{C^{5+}}$ in the non-ICME ACE/SWICS 1.1 data, plotted on a 10-base logarithmic scale. These ionization state ratios are used for the classification of solar wind because they provide insights into the temperature and ionization processes occurring in the solar corona. On top of it, the black lines represent the thresholds adopted to separate the data into solar wind types.

- The CH wind is separated from the slow Streamer wind using the threshold $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}} \leq 0.01$ (Von Steiger & Zurbuchen, 2016), displayed by the black solid line, and proton speed $V \geq 500$ km s^{-1}. Von Steiger and Zurbuchen (2016) have demonstrated from Ulysses data that the product of $\frac{O^{7+}}{O^{6+}}$ and $\frac{C^{6+}}{C^{5+}}$ is a better parameter to separate streamer from CH wind, than $\frac{O^{7+}}{O^{6+}}$ versus proton speed used for the classification of solar wind type released in the ACE/SWICS 1.1 database.
- The Outlier solar wind (Zhao, Landi, Lepri, Kocher, et al., 2017) is a subset of slow solar wind characterized by a decrease in the abundance of fully charged ions (e.g., He^{2+}, C^{6+}, O^{8+}) and lower ion temperatures compared to average Streamer solar wind distributions. Although this wind is slower and denser than Streamer slow wind, it seems to be accelerated from the same coronal source regions as the CH wind, for example, active regions and quiet-Sun regions, and its occurrence rate depends on the solar cycle. To isolate these Outliers, we take the data lying below the dot-dashed line, given by the empirical formula:

$$\log_{10}\left(\frac{C^{6+}}{C^{5+}}\right) = a_1 \log_{10}\left(\frac{O^{7+}}{O^{6+}}\right) + b_1 - c_1, \tag{5}$$

where $a_1 = 0.765$, $b_1 = 0.649$, and $c_1 = 0.602$ (see Equation 2 from Zhao, Landi, Lepri, Kocher, et al., 2017, for further details).

- The Upper Depleted Wind (UDW; Zhao et al., 2022) exhibits a systematic depletion of $\frac{O^{7+}}{O^{6+}}$. It is likely to be associated with quiet Sun regions, and its occurrence rate is anti-correlated with the solar cycle. To isolate the

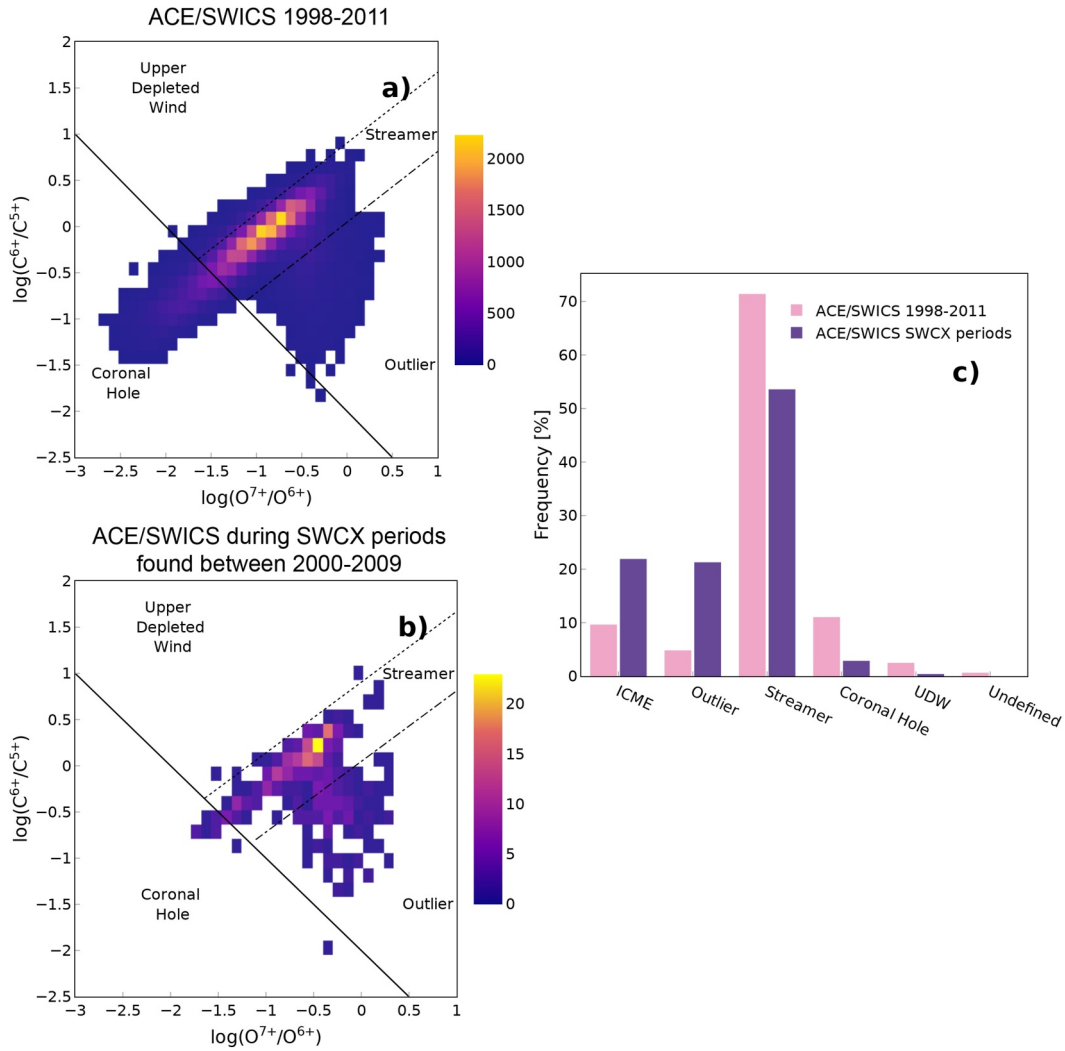

Figure 1. (a) 2D histogram of the $\log(C^{6+}/C^{5+})$ versus the $\log(O^{7+}/O^{6+})$ ratio counts from the non-ICME (Interplanetary Coronal Mass Ejection) ACE/SWICS 1.1 data. The black lines represent the Coronal Hole - Streamer (solid), the Outlier (dot-dashed) and the Upper Depleted Wind (UDW, dotted) type separation. Adapted from Koutroumpa (2024, Figure 3). (b) Same as (a) but with ACE/SWICS 1.1 data during the SWCX periods from Carter et al. (2011), which were found between 2000 and 2009. (c) Comparison of the solar wind type frequency, in percentage, for ACE SWICS 1.1 data during 1998–2011 (pink) versus SWCX periods (purple) found between 2000 and 2009.

UDW, we take the data lying above the dotted line, given by Equation 5 with $c_1 = -0.256$ (see Equation 2 from Zhao et al., 2022, for further details).

In cases where ACE/SWICS data is missing, the associated solar wind type for those timestamps is considered "undefined."

In Figure 1b, we apply the same set of rules as in Figure 1a to classify solar wind types, but we limit the ACE/SWICS 1.1 data set only to the time when geocoronal SWCX events were observed. Each SWCX event from the XMM SWCX data set has start and stop times, which indicate the duration of the corresponding XMM/EPIC exposure. The exposure length across the data set can range from a minimum of 2 hr to a maximum of one day, with an average duration of 9 hr. As in Figure 1, the Streamer category exhibits the highest occurrence, but the second most prevalent type of solar wind is Outlier, rather than CH.

In Figure 1c, we compare the occurrence frequency, in percentage, of each solar wind type for the ACE/SWICS measurements from 1998 to 2011 (in pink) and during the SWCX periods (in purple). Streamer slow wind is the

most common during both periods, with 71% (40,997 measurements) between 1998 and 2011 and 53% (262 measurements) during the SWCX periods. The frequency of CH solar wind is around 10% in the ACE/SWICS data set from 1998 to 2011, whereas, during SWCX periods, it drops to 3%. A lower percentage of CH wind being responsible for SWCX emissions was expected, due to lower ionization temperatures of their sources on the Sun, causing the elemental composition of the CH wind to be less fractionated (Neugebauer, 1994; Zhao, Landi, Lepri, Gilbert, et al., 2017).

The incidence of ICMEs increases from approximately 10% in the ACE/SWICS data set to 20% during SWCX events. ICMEs exhibit a higher $\frac{O^{7+}}{O^{6+}}$ ratio compared to the ambient solar wind (Richardson & Cane, 2004; Zhao et al., 2009). Upon reaching Earth, an ICME can temporarily distort the Earth's magnetosphere, pushing it Earthward. This phenomenon can expose a broader portion of the exosphere to the incoming solar wind. As the exosphere density rises sharply $\left(\propto r^3\right)$ with decreasing distance from Earth, an Earthward motion of the magnetopause result in sharp changes to SWCX emission (Samsonov et al., 2022). The increased frequency of ICMEs during SWCX events can be explained by the compound factors of the density of exospheric neutrals being proportional to SWCX emission (Cravens et al., 2001), and the dominance of the line emission from OVIII in most recorded SWCX events driven by ICMEs (Carter et al., 2011), due to charge state distributions shifted to higher charge states (Henke et al., 2001).

Unexpected to us is the higher frequency of the Outlier solar wind during SWCX periods, which reaches 20% of occurrence, against the 5% over all 5 categories in the period 1998–2011. While bare ions are more effective for the SWCX process (Bodewits et al., 2007), they were found to be less abundant in the Outlier wind compared to the Streamer wind (Zhao, Landi, Lepri, Kocher, et al., 2017), suggesting that we might have more He-like ions enhancing the SWCX emission during such periods.

To assess whether the distribution of solar wind types in the ACE 1998–2011 data is statistically different from the distribution during the SWCX periods, we use the chi-square test for independence. The test yields a chi-square statistic of 401.47 with 5 degrees of freedom, and a p-value $\ll 10^{-5}$, suggesting that the solar wind sampled during SWCX periods is not a random subset of the parent ACE data set from 1998 to 2011.

3.2. ACE Proton Properties of the Solar Wind Driving SWCX

We investigate the proton properties of the solar wind during SWCX periods. In Figure 2, we depict the proton speed (a), density (c), flux (e), and temperature (g) measured by ACE/SWICS and ACE/SWEPAM during non-ICME SWCX events. Each solar wind property is shown by 2D histograms in the left panel, divided into 20 × 20 boxes, with each box indicating the solar wind property value averaged over their occurrence between 2000 and 2009 (shown in Figure 1b). The black lines indicate the thresholds used for identifying non-ICME solar wind types, as detailed in Sect. 3.1. Panels (b), (d), (f), and (h) show histograms of the occurrence rate for Streamer (black), Outlier (red), and ICME (green) solar winds, plotted on a 10-base logarithmic scale. Table 1 presents a summary of the average solar wind properties along with the IMF from ACE/MAG during SWCX periods, compared to the averages recorded from 1998 to 2011 for the different types of solar wind.

Figures 2a and 2b reveals similar proton speed distributions for the Streamer and Outlier winds, with an average value of 380.3 km s^{-1} and 387.6 km s^{-1} respectively. In Figure 2a, speed tends to increase to the lower left of the diagram, as also shown by Von Steiger and Zurbuchen (2016). Figure 2b shows, along with Outlier and Streamer proton speed histograms, the occurrence rate of the ICME proton speed during the SWCX periods, peaking at 400 km s^{-1} with an average value of about 450 km s^{-1}. The values from Table 1 do not indicate significant changes between the average proton speed during the SWCX periods and 1998–2011.

Figures 2c and 2d display the solar wind proton density distribution. Figure 2c shows distinct density patterns between the Outlier and Streamer wind, with the Outlier wind exhibiting higher density values compared to the Streamer wind. Figure 2d shows the Streamer wind histogram centered around 7 cm^{-3}, while the Outlier wind peaks at 11.7 cm^{-3}, consistent with values previously reported in Zhao, Landi, Lepri, Kocher, et al. (2017). According to Table 1, the mean Streamer density is 13.5 cm^{-3}, double the average density from 1998 to 2011. This results in a drastic reduction in the relative difference between the Outlier and Streamer proton density from 100.9% in 1998–2011 to 8.9% during the SWCX periods. This trend suggests, as one might expect, that SWCX events are most likely to be observed when the solar wind density is high.

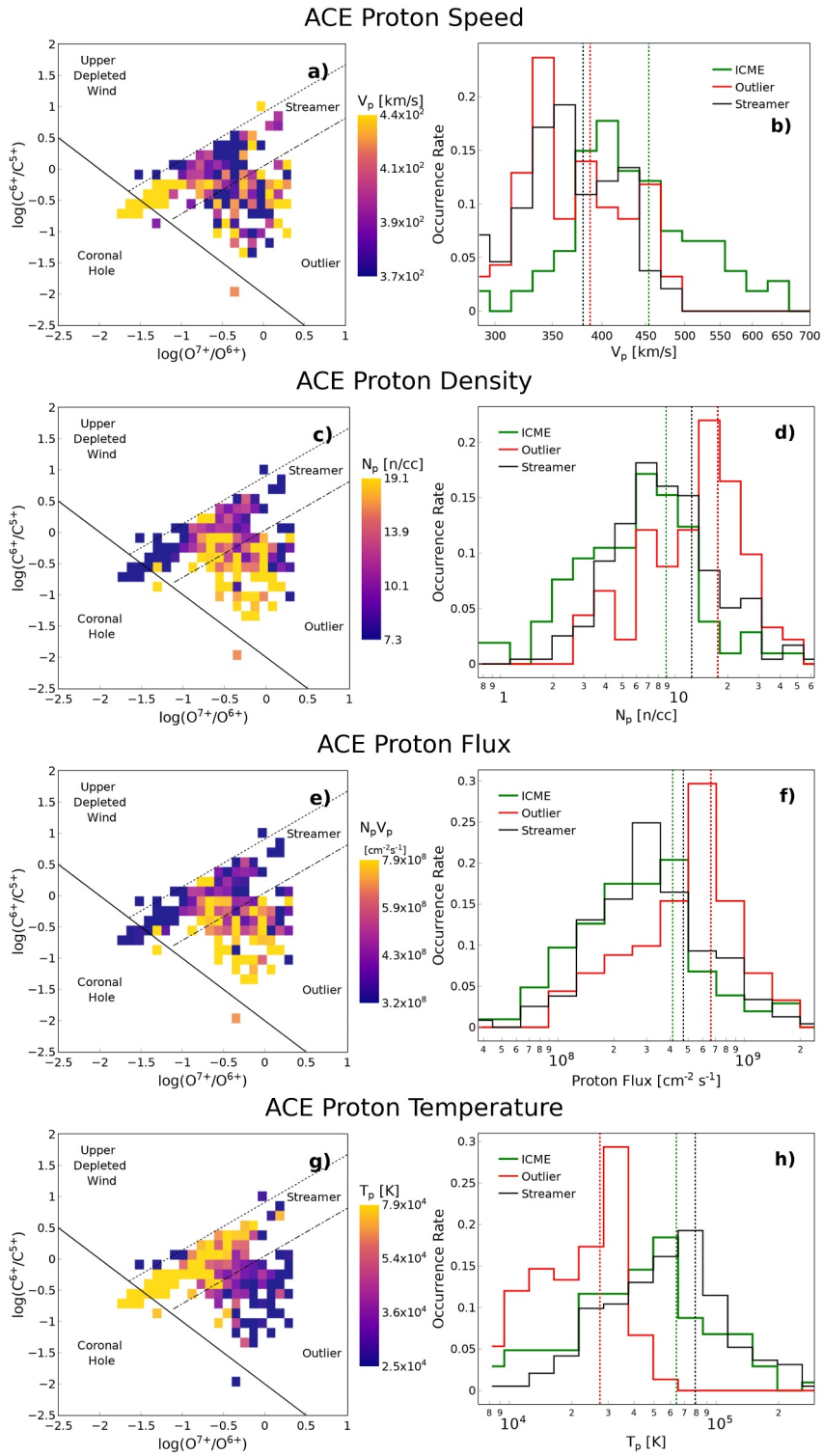

Figure 2. Distributions of proton speed (V_p), density (N_p), flux ($N_p \times V_p$), temperature (T_p) and interplanetary magnetic field during non-ICME driven SWCX events. In the left panels, (a), (c), (e), (g), (i), the solar wind properties are shown as 2D histograms, averaged over their occurrence between 2000 and 2009, in a $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}}$ space, with black lines separating different solar wind types. The right panels, (b), (d), (f), (h), (j), show 1D histograms that indicate the occurrence rate of solar wind properties for the Streamer (black) and Outlier (red) wind.

Table 1
Comparison of Averages, Standard Deviations, and Percentage Changes of the Proton Speed, Density, Flux, and Magnetic Field of the Outlier Slow Wind and Streamer Slow Wind Measurements During SWCX Events Versus the Values for the ACE Data Set From 1998 to 2011, With and Without Interplanetary Coronal Mass Ejection Periods

		Solar wind types			Relative difference	
Mean (std)	Data set	Streamer	Outlier	ICME	$\frac{\text{Outlier}-\text{Streamer}}{\text{Streamer}}$ %	$\frac{\text{ICME}-\text{Streamer}}{\text{Streamer}}$ %
Proton Speed [km/s]	SWCX Periods	380.3 (48.3)	387.6 (51.6)	454.0 (75.7)	1.9	19.4
	1998–2011	389.6 (56.7)	372.1 (55.7)	466.0 (111.6)	−4.51	19.6
Proton Density [n/cc]	SWCX Periods	13.5 (9.84)	14.73 (9.23)	8.54 (6.21)	8.89	−36.9
	1998–2011	6.41 (4.51)	12.88 (8.69)	7.12 (7.03)	100.93	11.05
Proton flux (10^8 cm^{-2}s^{-1})	SWCX Periods	4.71 (3.70)	6.62 (3.86)	4.13 (3.92)	40.5	−12.42
	1998–2011	2.60 (1.84)	4.96 (3.29)	3.29 (3.36)	90.84	26.52
Proton temperature (10^4 K)	SWCX Periods	7.91 (5.47)	2.74 (1.24)	6.41 (5.05)	−65.4	−18.95
	1998–2011	7.60 (4.33)	2.62 (1.46)	9.55 (10.6)	−65.6	25.53
IMF [nT]	SWCX Periods	8.63 (4.41)	8.69 (4.27)	10.77 (4.97)	0.70	24.73
	1998–2011	5.47 (2.26)	7.09 (3.91)	8.87 (5.51)	29.63	62.1

In Figures 2e and 2f, the proton flux, calculated as the product of the proton speed and the proton density, resembles the patterns observed in the proton density. The average Outlier wind flux peaks at 6×10^8 cm^{-2} s^{-1}, and is higher than the Streamer wind flux by 40.5%. Since the SWCX emission is proportional to the solar wind flux, the higher incidence of Outlier wind during SWCX periods may be explained by associated higher proton flux during such wind type. The ICME wind clipped during SWCX periods shows a similar proton flux distribution compared to the Streamer wind, with a relative deviation of −12% (Table 1), and an average value of 4.1×10^8 cm^{-2} s^{-1}.

Figures 2g and 2h show the proton temperature distribution. In Figure 2g, a clear separation between the Streamer wind and the Outlier wind can be observed according to the proton temperature, with the Outlier wind having a significantly lower temperature than the Streamer wind by 65%. This difference between the average values is consistent with the period 1998–2011. The ICME wind, on the other hand, has a temperature 18% lower than the Streamer temperature, on average, and also decreases by approximately 30% compared to 1998–2011. Thus, we see no strong dependence of SWCX events on the temperature of the solar wind.

The results for the IMF distributions are summarized together with the other properties in Table 1. The IMF averages during Streamer and Outlier wind almost match around 8.6 nT, with a relative difference between the Outlier and Streamer wind of only 0.7%. This difference is about 30% between 1998 and 2011. The ICME IMF increases from 8.9 nT during 1998–2011 to 10.8 nT during the SWCX periods.

3.3. XMM Lightcurve Variability by Solar Wind Type

The XMM lightcurves are a timeseries of soft X-ray photon counts binned at 1 kilosecond (ks) intervals, between 500 and 700 eV, during periods of detected SWCX. These lightcurves are already filtered by Carter et al. (2011) both temporally, to remove periods of soft proton flares, and spatially, for bright sources in the cameras' FOV. Figure 3a shows as an example the XMM lightcurve in counts per kilosecond (ct ks^{-1}) of the observation with ID 0111300101 during revolution 0340. The red horizontal line is the mean count rate during that observation and the arrows departing from the mean represent the deviation of the count rate from the mean, positive when directed upward, and negative otherwise. The inflowing solar wind type is shown in the plot by different background colors.

To analyze whether XMM is able to distinguish the examined solar wind types, we look at the variability in the XMM lightcurve during our list of SWCX periods. We measure the lightcurve variability as the deviation of the XMM count rate from the mean of each observation, then averaged with a 2 hr cadence (~7 ks). In such a way, the timeseries of the lightcurve deviation are synchronous with ACE/SWICS data and short-term fluctuations of the count rate around the mean are canceled out. In Figure 3b, the $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}}$ chart is populated with the XMM

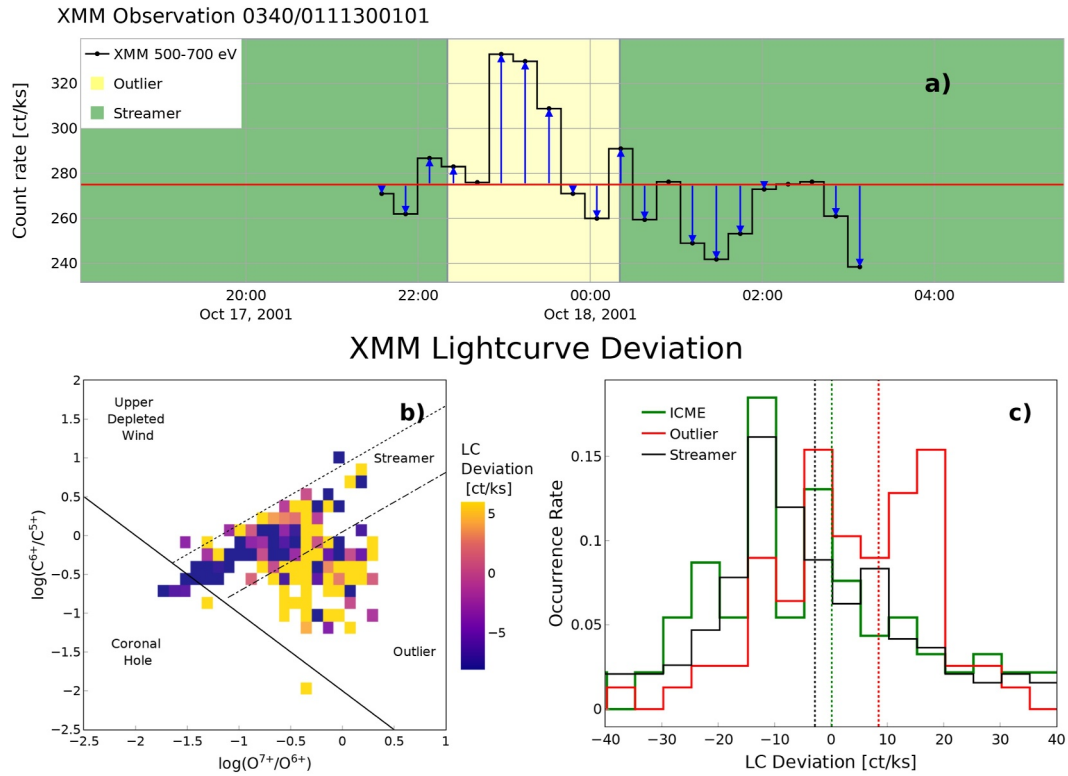

Figure 3. (a) In black, the XMM lightcurve in ct/ks, between 500 and 700 eV, of the observation with ID 0111300101 during revolution 0340. The red horizontal line is the mean count rate during that observation and the arrows departing from the mean represent the deviation of the count rate from the mean, positive when directed upward, and negative otherwise. The inflowing solar wind type is shown in the plot by different background colors. (b) 2D histogram of the XMM lightcurve deviation averaged over their occurrence between 2000 and 2009, in the $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}}$ space, with black lines separating different solar wind types. (c) Occurrence rate of XMM lightcurve deviation for the Streamer (black), Outlier (red), and interplanetary coronal mass ejection (green) wind.

lightcurve deviation data, where each box represents the average lightcurve deviation based on the corresponding $\frac{O^{7+}}{O^{6+}}$ and $\frac{C^{6+}}{C^{5+}}$ values over time from the ACE data. The plot reveals a higher number of negative values within the Streamer region compared to the Outlier region, where the values are predominantly positive with only a few boxes showing a lightcurve deviation below 0 ct/ks. Thus, the Outlier period has stronger SWCX emission than the Streamer periods, as seen by multiple different visualizations of the same data.

Figure 3c displays the occurrence rate distributions of lightcurve deviation for Outlier, Streamer, and ICME solar wind types, shown in red, black, and green respectively. The ICME and Streamer distributions cover the same range of [−40, +40] ct/ks and are skewed toward negative values, with peaks at approximately −20 ct/ks and −12 ct/ks respectively. Despite this skew, their means are close to 0: at −3 ct/ks for ICME and −1 ct/ks for Streamer. The Outlier distribution, depicted in red, is the only distribution with a mean in the positive values at 9 ct/ks. Moreover, the Outlier occurrence rate distribution features two peaks of equal height, one around 0 ct/ks and another at 19 ct/ks.

3.4. ACE Heavy Ion Abundance Versus XMM Ion Line Fluxes

We conduct a comparison between individual heavy ion abundances from ACE and heavy ion line fluxes obtained through spectral analysis of the XMM SWCX data set. The ACE abundances are considered relative to protons, allowing us to filter out any variations in heavy ion abundance resulting from changes in solar wind density, and consequently, proton density. As for the SWCX data set by Carter et al. (2011), it is crucial to note that it presents only constant values of heavy ion line fluxes throughout the duration of each SWCX-affected XMM observation; there is no way to normalize these observations to the proton flux. These values correspond to the average fluxes

Table 2
Comparison of Averages, Standard Deviations, and Percentage Changes of the XMM Ion Fluxes and ACE Ion Abundances, During SWCX Periods Versus Averages for the ACE Data Set From 1998 to 2011

Mean (std)	Data set	Solar wind types			Relative difference	
		Streamer	Outlier	ICME	$\frac{\text{Outlier}-\text{Streamer}}{\text{Streamer}}$ %	$\frac{\text{ICME}-\text{Streamer}}{\text{Streamer}}$ %
O^{8+}/p abundance $[10^{-5}]$	1998–2011	0.37 (0.85)	1.01 (1.71)	0.52 (0.76)	170.6	546.2
	SWCX Periods	1.09 (1.80)	1.70 (1.83)	4.53 (4.93)	56.8	316.9
OVIII line flux [10^{-13} ergs cm^{-2} s^{-1}]	SWCX Periods	0.90 (0.85)	1.38 (1.01)	2.10 (2.04)	52.9	133.1
O^{7+}/p abundance $[10^{-5}]$	1998–2011	4.33 (4.11)	56.9 (113.3)	9.33 (18.9)	1,214.1	115.5
	SWCX Periods	8.58 (7.21)	98.0 (137.1)	22.8 (74.4)	1,043.0	165.7
OVII line flux [10^{-13} ergs cm^{-2} s^{-1}]	SWCX Periods	2.14 (1.52)	1.78 (1.47)	1.57 (1.55)	−16.2	−26.1
C^{6+}/p abundance $[10^{-5}]$	1998–2011	7.33 (4.91)	5.82 (5.33)	9.71 (8.04)	−20.6	32.5
	SWCX Periods	9.76 (6.22)	7.84 (5.46)	14.2 (10.6)	−19.6	45.5
CVI line flux [10^{-13} ergs cm^{-2} s^{-1}]	SWCX Periods	1.09 (1.35)	1.03 (1.04)	1.46 (1.58)	−5.96	33.5
C^{5+}/p abundance $[10^{-5}]$	1998–2011	7.99 (4.52)	30.8 (52.5)	7.88 (11.2)	285.0	−1.27
	SWCX Periods	8.64 (5.25)	41.1 (52.8)	10.1 (14.2)	373.7	16.7
CV line flux [10^{-13} ergs cm^{-2} s^{-1}]	SWCX Periods	3.07 (2.85)	2.76 (3.01)	2.94 (3.62)	−9.97	−5.46
Mg^{11+}/p abundance $[10^{-6}]$	1998–2011	0.13 (0.38)	6.13 (31.2)	2.06 (16.7)	4,394.9	1414.5
	SWCX Periods	0.34 (0.89)	9.24 (13.9)	2.99 (10.5)	2,624.8	783.8
MgXI line flux [10^{-14} ergs cm^{-2} s^{-1}]	SWCX Periods	1.23 (1.64)	1.83 (2.45)	2.85 (5.62)	47.89	130.6

extracted during each SWCX event. Since there is no direct way to cross-calibrate ion abundances from ACE and ion line fluxes from XMM spectral analysis we plot ACE and XMM data separately and compare their distributions and relative differences across the solar wind types, summarized in Table 2.

In Figure 4, we compare the distributions of the bare oxygen abundance, O^{8+}/p, from ACE, and the emitted line flux, OVIII, from XMM. Figures 4a and 4c respectively show higher values of O^{8+}/p and OVIII for higher $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}}$. As a result, in the Outlier wind, O^{8+}/p is on average 57% more abundant compared to the Streamer wind, and the OVIII line flux is stronger by 53% (from Table 2). We are aware that the results presented in Table 2 differ from those in Zhao, Landi, Lepri, Kocher, et al. (2017) (Table 3), which report a depletion of fully stripped ions, such as O^{8+}, in the Outlier wind compared to the Streamer wind. We believe this discrepancy is caused by the different versions of the ACE data set being used, with ours being publicly available from CDAWeb and Zhao's the pre-distributed data set available to the instrument team. This discrepancy, however, does not affect the results of this analysis, which compare the variability observed in ACE data with that of XMM. The O^{8+}/p distribution for ICME, illustrated in Figure 4b, is skewed toward higher values, with an average abundance 133% greater than in the Streamer wind. This elevated O^{8+}/p abundance is characteristic of ICME wind composition, confirming expected results based on literature findings (Richardson & Cane, 2004).

Figure 5a displays a clear separation between the Outlier and Streamer winds in terms of O^{7+}/p, with the Outlier wind showing a greater abundance compared to the Streamer wind. This separation becomes more pronounced in Figure 5b, where the occurrence rate distributions for the Outlier and Streamer winds peak at approximately 5×10^{-4} and 5×10^{-3}, respectively, indicating an order of magnitude difference. The distribution of O^{7+}/p in the ICME wind lies between these two, overlapping mainly with the Streamer distribution, and peaks at 1.6×10^{-4}. In Figure 5c, the OVII line flux does *not* show a distribution similar to O^{7+}/p, and no clear pattern can be observed. The occurrence rate distribution for ICME, Streamer, and Outlier winds seems to overlap, with the main peak at 1.8×10^{-13} ergs cm^{-2} s^{-1}. However, the ICME and Outlier distributions exhibit a second peak at 4×10^{-12} ergs cm^{-2} s^{-1}, which is not observed in the Streamer distribution. The combination of Figures 4 and 5 can be understood as follows: SWCX periods were determined based on times when the 0.5–0.7 keV band, containing both OVII and OVIII, is strong. Between the two lines, OVII is the stronger one. Therefore, SWCX event periods are generally selected when OVII is high (as seen in Figure 5c), which explains why there are fewer

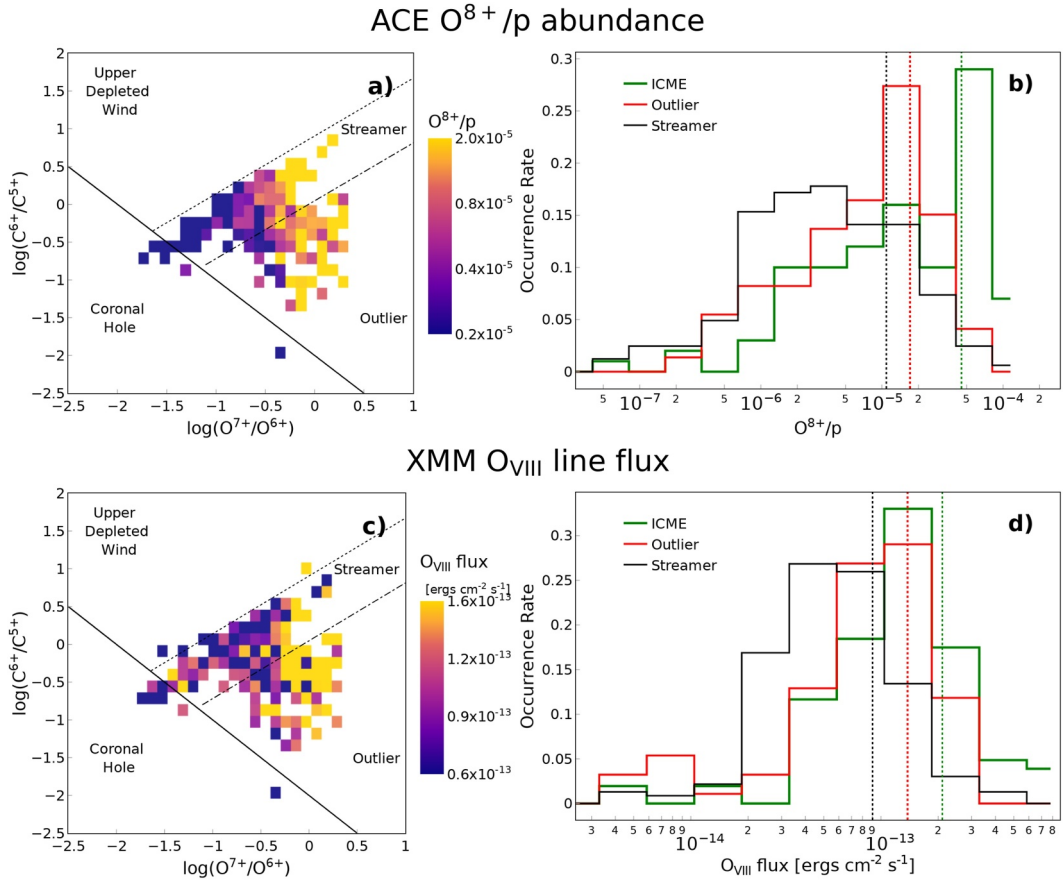

Figure 4. (a) SWCX periods ACE O^{8+}/p 2D histograms, averaged over their occurrence between 2000 and 2009, in a $\frac{O^{7+}}{O^{6+}} \times \frac{C^{6+}}{C^{5+}}$ space, with black lines separating different solar wind types. (b). Occurrence rate of ACE O^{8+}/p for the Streamer (black) and Outlier (red) wind. (c) (d) Same as left plots but for OVIII ion line fluxes from the XMM SWCX data set.

CH wind periods associated with SWCX. After selecting periods with high OVII emission, whether OVIII is also high depends upon the type of solar wind.

Figure 6a illustrates a greater abundance of C^{6+}/p at the upper edge of the Streamer wind, and a depletion of C^{6+}/p in the Outlier wind compared to the Streamer wind, which is consistent with Figure 6e of Zhao, Landi, Lepri, Kocher, et al. (2017). However, the occurrence rate distributions of Streamer, ICME, and Outlier winds span the same range of C^{6+}/p values, all peaking at approximately 8×10^{-5}. Similarly, the CVI line flux shows

Table 3
Performance Metrics With and Without XMM Features (Including Standard Deviations)

		Labels			Averaged metrics	
Metrics	Features	Streamer	Outlier	ICME	Macro avg	Accuracy
Precision	ACE + XMM	0.86 (0.06)	0.76 (0.12)	0.84 (0.07)	0.82 (0.06)	0.83 (0.05)
Recall		0.90 (0.07)	0.61 (0.14)	0.87 (0.07)	0.79 (0.06)	
F1 score		0.88 (0.05)	0.67 (0.12)	0.85 (0.05)	0.80 (0.06)	
Precision	ACE	0.46 (0.38)	0.76 (0.26)	0.49 (0.33)	0.57 (0.12)	0.74 (0.16)
Recall		0.35 (0.29)	0.82 (0.17)	0.54 (0.30)	0.57 (0.13)	
F1 score		0.40 (0.33)	0.76 (0.20)	0.50 (0.30)	0.55 (0.12)	

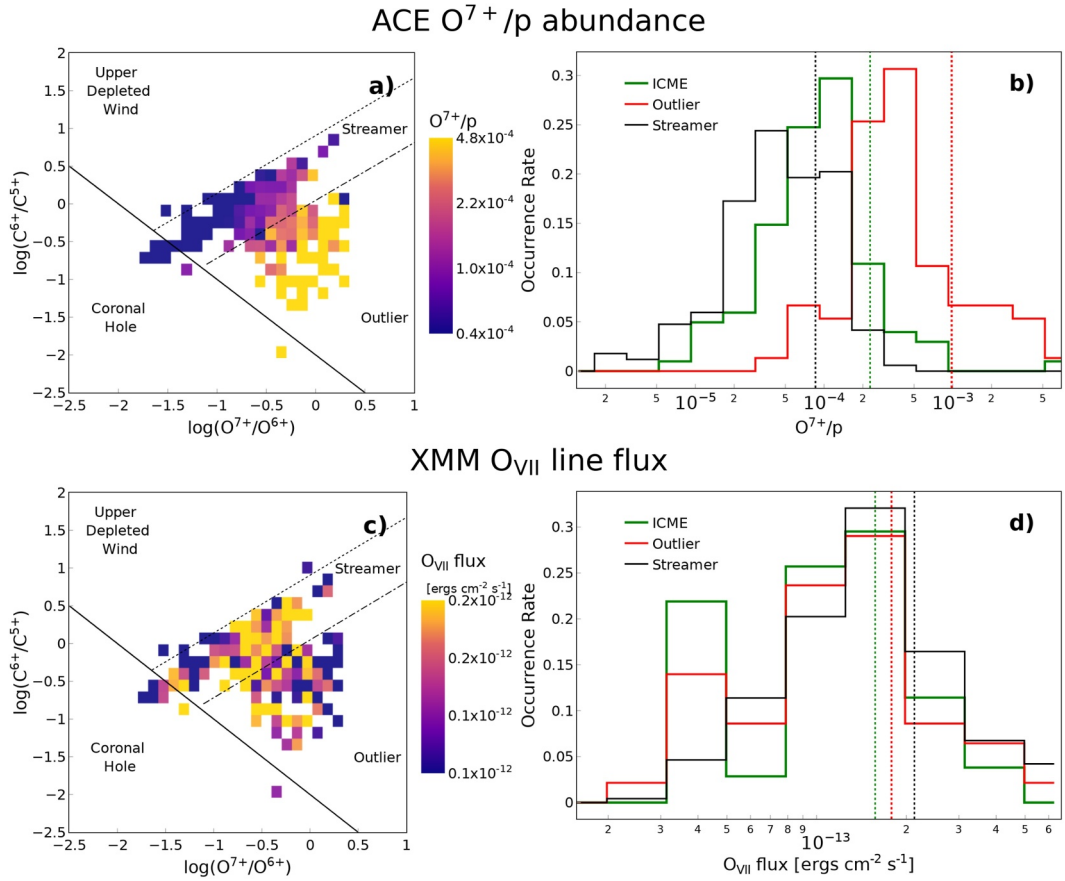

Figure 5. Same as Figure 4 but for ACE O^{7+}/p on (a) and (b), and OVII ion line fluxes from the XMM SWCX data set on (c) and (d).

limited differentiation between the solar wind types, indicating that neither C^{6+}/p nor CVI are not effective indicators for distinguishing between different solar wind types during SWCX periods.

The variability observed in C^{5+}/p across the solar wind types is completely missed in CV line flux, from Figures 7c and 7d, which show nearly overlapping distributions for ICME, Outlier, and Streamer winds.

In Figure 8, Mg^{11+}/p is higher as $\frac{O^{7+}}{O^{6+}}$ increases. The average Mg^{11+}/p in the Streamer wind is about an order of magnitude lower than in the Outlier and ICME wind. However, the MgXI line flux distribution across the three solar wind types highly overlap, with a similar peak at around 2×10^{-14} ergs cm^{-2} s^{-1}.

From Table 2, the ACE average ion abundances relative to proton during SWCX periods are to some extent always higher than in 1998–2011, showing that the time-variable SWCX process tends to happen during richer solar winds.

4. Random Forest Classifier for Solar Wind Type Recognition

In this section, we aim to see whether the aid of XMM-Newton data can improve predictions of solar wind type, in the absence of heavy ion measurements from ACE/SWICS. Our data set comprises a combination of solar wind proton properties from ACE/SWEPAM (proton density, velocity, flux, and temperature), IMF from ACE/MAG and XMM-derived features (lightcurve deviation and line fluxes from OVIII, OVII, CVI, CV and MgXI), as described in previous sections. In addition, we include two temporal features, the day of year (DOY) of the measurement and the corresponding monthly averaged sunspot number from NOAA (https://www.swpc.noaa.gov/products/solar-cycle-progression). The solar wind types that we aim to predict are Streamer, Outlier, and ICME wind, as they are the predominant types associated with SWCX.

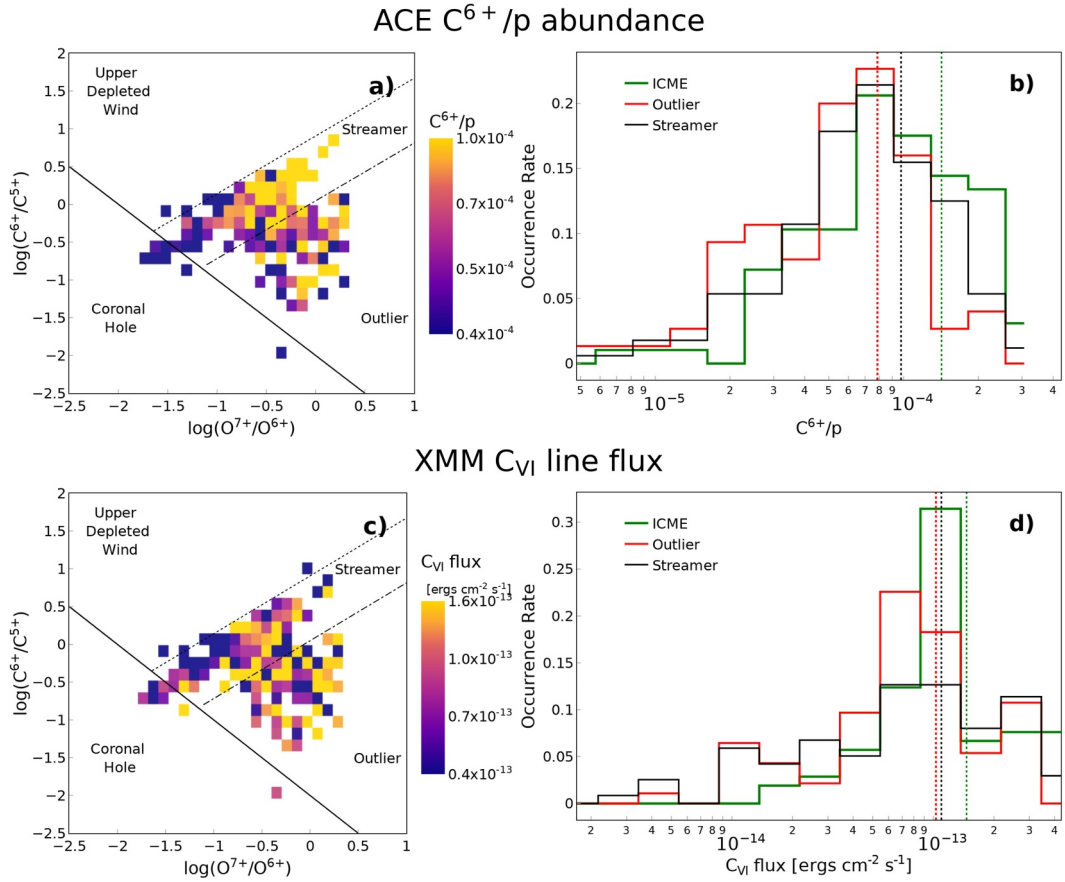

Figure 6. Same as Figure 4 but for ACE C^{6+}/p on (a) and (b), and CVI ion line fluxes from the XMM SWCX data set on (c) and (d).

We predict solar wind types employing a Random Forest (Breiman, 2001) classifier from the *scikit-learn* library for Python (Pedregosa et al., 2011). Random Forest is a supervised ensemble learning method that constructs multiple decision trees during training and merges their results to improve classification performance. One of the advantages of using Random Forests is inherently performed feature selection. During the construction of each tree within the forest, the model evaluates the importance of each feature based on how well it improves the purity of the node splits. The final prediction is made by aggregating the predictions from all individual trees, through majority voting for classification. This approach helps to reduce over-fitting and improve the model's generalization to unseen data.

To ensure that the features were on a comparable scale, we standardized them, so that each feature has a mean of 0 and a standard deviation of 1. This normalization process is usually not necessary for the performance of Random Forests, but we found it helpful given the mixture of data types and different scales in our data set. The hyperparameters of the Random Forest model are tuned using *RandomizedSearchCV* (https://scikit-learn.org/1.5/modules/generated/sklearn.model_selection.RandomizedSearchCV.html). This method searches for the hyperparameters that maximize the model accuracy, by randomly sampling from a specified distribution of possible parameter values. The final parameters for the model are set as follows: bootstrap was set to False, maximum depth of the tree set to 10, minimum number of samples required to split an internal node to 2, minimum number of samples required to be at a leaf node to 1, and number of trees in the forest to 100.

To evaluate the performance of the model, we employ a cross-validation approach using *RepeatedStratifiedKFold* (https://scikit-learn.org/1.5/modules/generated/sklearn.model_selection.RepeatedStratifiedKFold.html) with 5 splits and 3 repeats. The process involves splitting the data set into training and validation sets, training the Random Forest model on the training set, and evaluating it on the validation set. We chose this method because it ensures that each fold maintains approximately the same proportion of samples from each target class as the full

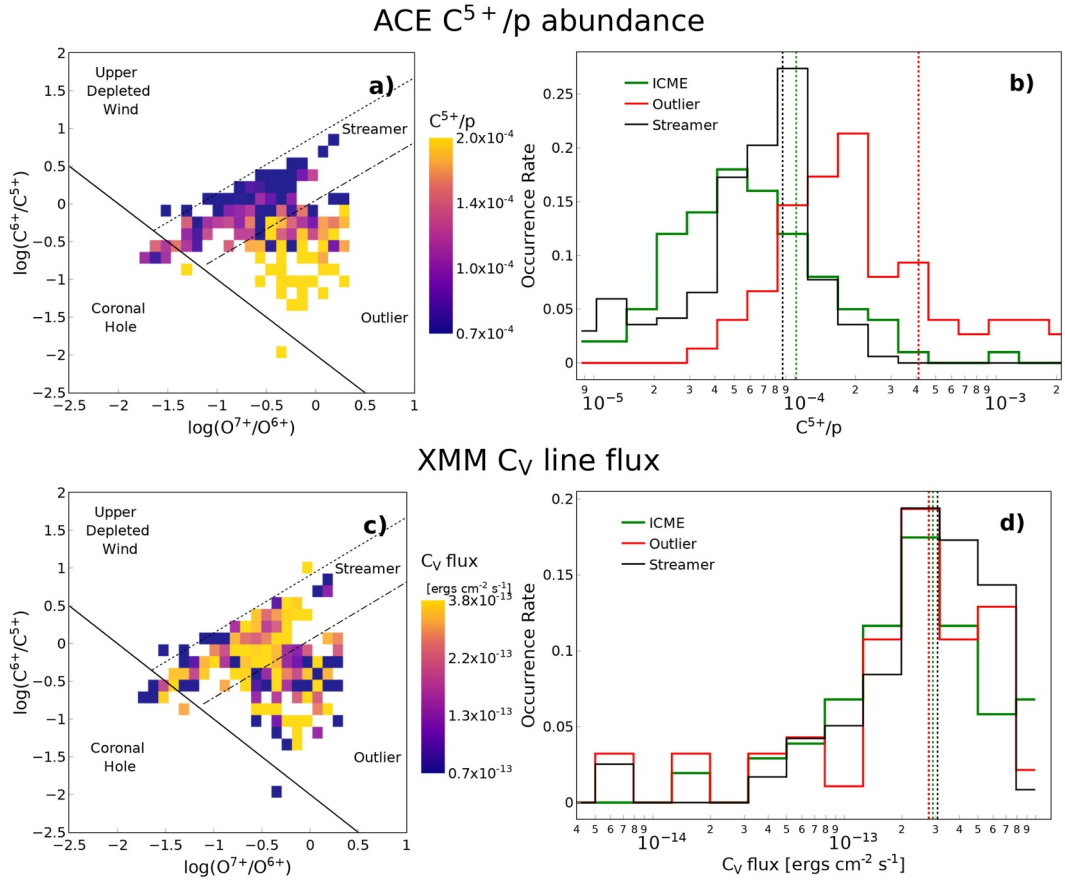

Figure 7. Same as Figure 4 but for ACE C^{5+}/p on (a) and (b), and CV ion line fluxes from the XMM SWCX data set on (c) and (d).

data set, preserving class balance across the splits. Performance metrics are presented in Table 3 for models trained with and without the inclusion of XMM features. The metrics used include precision, recall, and F1 scores, which are calculated for each fold and then averaged. Precision measures the ratio of correctly predicted positive observations to the total predicted positives, while recall measures the ratio of correctly predicted positive observations to the total actual positives. The F1 score,

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{6}$$

is the harmonic mean of precision and recall, providing a single metric that balances both measures. The macro-averaged metrics compute the average precision, recall, and F1 score across all classes, with all classes equally contributing to the final averaged metric. This metric is useful in case of a multi-class classification problem with imbalanced classes, like the one we are analyzing, because it will highlight the performance of a model on all classes equally. Finally, accuracy measures the overall correct predictions divided by the total number of predictions.

The comparison between models trained with and without XMM features shows significant improvements in most metrics. Notably, the macro-averaged precision, recall, and F1 score increase from 0.57 to 0.82, 0.57 to 0.79, and 0.55 to 0.80, respectively, when transitioning from ACE-only to ACE and XMM features. The overall accuracy also increases from 0.74 to 0.83. When looking at individual classes, the F1 score shows a substantial improvement for Streamer and ICME class prediction, although similar gains are not seen for the Outlier class. Overall, the model trained with both ACE and XMM features achieves higher scores and exhibits much smaller standard deviations, suggesting enhanced stability.

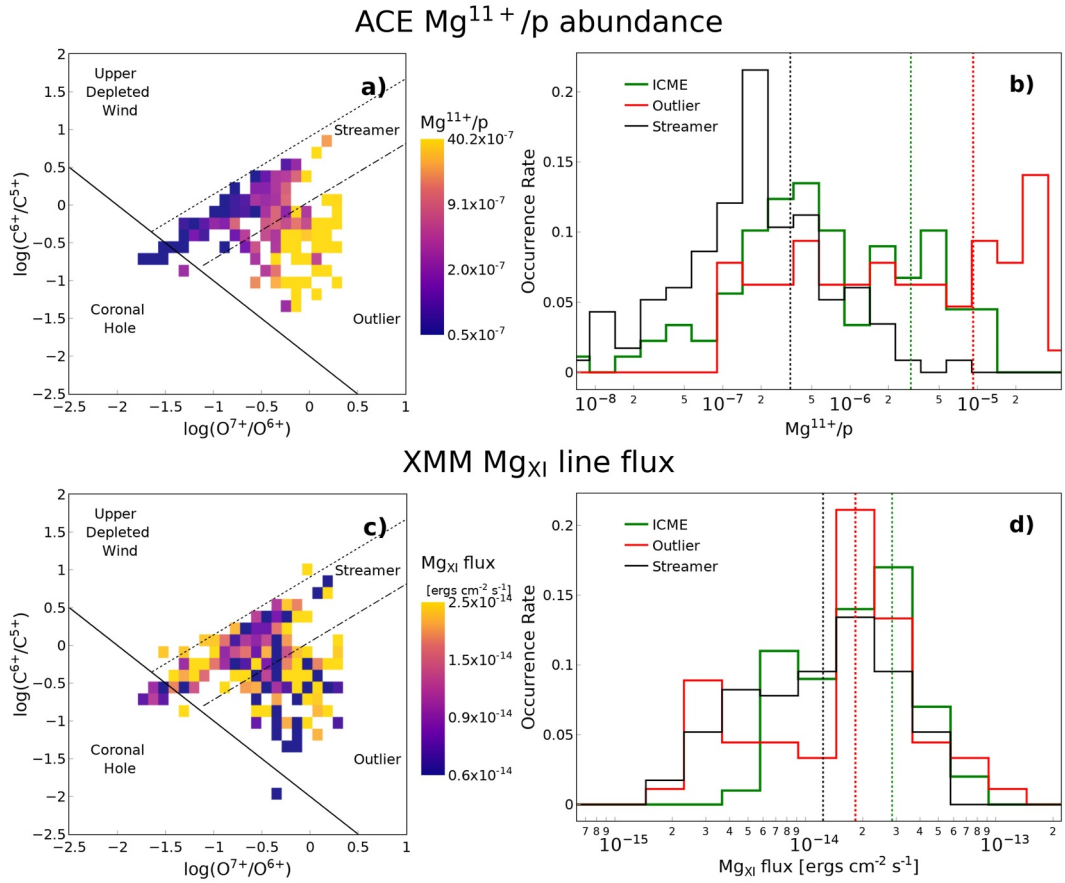

Figure 8. Same as Figure 4 but for ACE Mg^{11+}/p on (a) and (b), and MgXI ion line fluxes from the XMM SWCX data set on (c) and (d).

Figure 9a displays the average confusion matrix for the model trained with ACE and XMM features, providing a visual representation of classification performance averaged across all folds and repeats, with standard deviations in brackets. This matrix details the counts of true positive, true negative, false positive, and false negative predictions across all classes. The *x*-axis represents the predicted solar wind type, while the *y*-axis represents the true solar wind type. On average, the test set contains 47 samples: 30 from the Streamer wind, 11 from the Outlier wind, and 16 from the ICME wind. The majority of samples from each solar wind type are correctly predicted, as

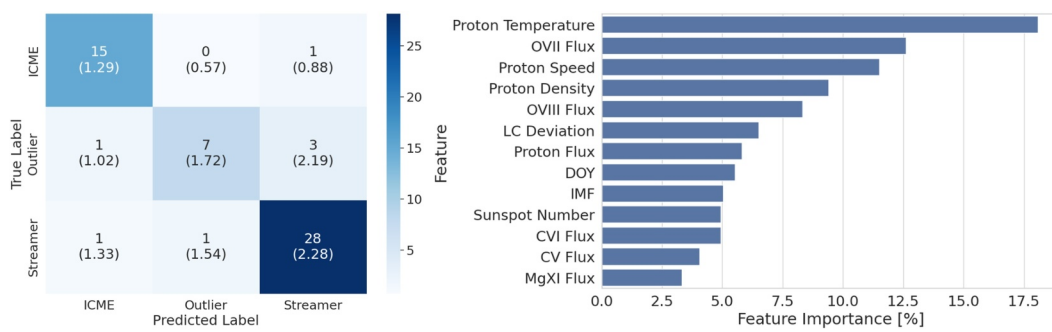

Figure 9. *Left:* Average confusion matrix for the Random Forest model trained with ACE and XMM features, with standard deviations in brackets. The *x*-axis corresponds to the predicted solar wind type, and the *y*-axis corresponds to the true solar wind type. The diagonal elements indicate the number of correct predictions for each class, while the off-diagonal elements show the misclassifications. *Right:* Importance of each feature in predicting solar wind types, expressed as a percentage. The features are ranked in descending order of their importance.

evidenced by the values along the diagonal. Specifically, there is an average of one ICME sample misclassified as Streamer per fold, and from the Streamer group, an average of 2 samples are misclassified as ICME and Outlier wind. As indicated in Table 3, the Outlier wind exhibits a lower performance compared to the other classes. The confusion matrix shows that missed Outlier samples are labeled as Streamer 3 out of 11 times on average, and as ICME 1 out of 11 times. This highlights the model's tendency to confuse Outlier samples with the Streamer class more frequently.

Figure 9b shows the feature importance analysis for the Random Forest model trained with ACE and XMM features, which highlights the most influential variables in predicting solar wind types. The top six features are half from ACE and half from XMM, showing how the addition of XMM features to train the Random Forest model contributes to the model's predictive accuracy. In order we have: proton temperature, OVII flux, and proton speed, proton density, OVIII flux and lightcurve deviation, with importance scores of 18.1%, 12.6%, and 11.5%, 9.4%, 8.3% and 6.5%. These findings are expected as many of these features exhibited the greatest relative separation between the different wind types in our preliminary analysis. Specifically, OVIII flux's high ranking is anticipated because it was the only feature displaying similar distributions to the parent ion abundance O^{8+} from ACE. Unexpectedly, OVII flux ranks in second place, despite its distribution largely overlapping across solar wind types in Figure 5d. This high ranking appears to be influenced primarily by its distribution along the *y*-axis for flux values below 8×10^{-14} ergs cm^{-2} s^{-1}. To investigate this effect, we excluded these lower flux values from the data set, which led to the OVII flux ranking dropping from second to sixth place, confirming that its initial high ranking is largely driven by these low flux values. In the process, the rankings of the other features were marginally affected.

In contrast, the XMM features, including CVI flux, CV flux, and MgXI flux, are ranked at the bottom. This aligns with our expectations since their distributions (Figures 6–8) showed limited to no separation between the different classes.

5. Conclusions

In this study, we analyzed data derived from 103 XMM-Newton observations exhibiting time-variable exospheric Solar Wind Charge Exchange (SWCX) emissions (Carter et al., 2011), alongside concurrent ACE data on heavy ions (e.g., C^{6+}, C^{5+}, O^{8+}), proton properties and IMF. We investigated the predominant solar wind types driving periods of geocoronal SWCX (SWCX periods) and analyzed the characteristics of their ion and proton properties. The solar wind types are obtained from literature, as in Koutroumpa (2024), and they are differentiated based on ion abundances and proton speed measurements from ACE. Overall, the SWCX periods display a solar wind type distribution distinct from the average ACE data from 1998 to 2011. Even though XMM-Newton measures charge exchange line emissions along a line-of-sight with moderate spectral resolution, unlike ACE's in situ measurements of parent ion abundances, XMM-Newton is able to capture distributions that deviate from the general population and can effectively identify solar wind types during SWCX events.

We observed that a subset of the slow solar wind, known as Outlier wind, characterized by systematically lower ion and proton temperatures compared to the average Streamer wind (Zhao, Landi, Lepri, Kocher, et al., 2017), occurs more frequently during SWCX periods, around 20%, than in the overall ACE data set from 1998 to 2011, about 5%. The Outlier wind exhibits higher solar wind flux compared to the Streamer wind, what we called the average slow solar wind, which could explain its increased occurrence during SWCX periods. Moreover, the Outlier is characterized by high OVII flux, as detected by spectral analysis of XMM-Newton EPIC data. Since the OVII line is the strongest of the batch, detection of that SWCX line is more probable than detecting the other lines.

The remaining SWCX periods are predominantly associated with Streamer wind, approximately 70% and interplanetary coronal mass ejections (ICMEs, as defined by Richardson & Cane, 2010), around 20%. Fast solar wind from CHs is almost absent during the SWCX periods we analyzed, probably due to their lower OVII (or O^{7+}) values. This was expected due to the lower temperatures of their sources on the Sun, which cause the elemental composition of the fast solar wind to be less fractionated (Neugebauer, 1994; Zhao, Landi, Lepri, Gilbert, et al., 2017).

We compared individual heavy ion abundances from ACE with heavy ion line fluxes obtained through spectral analysis of the XMM SWCX data set. It is crucial to note that XMM does not provide in situ measurements of X-ray

emission; instead, it integrates the observed emission along its line-of-sight based on the selected energy range. Consequently, there is no direct method for cross-calibrating ACE ion abundances with XMM spectral line fluxes. Therefore, we analyzed ACE and XMM data independently, comparing their distributions and relative differences across various types of incoming solar wind. We found that both the XMM data and ACE data reveal an increasing O^{8+}/p abundance and OVIII flux with a rising ratio of $\frac{O^{7+}}{O^{6+}}$ during the events from the SWCX data set. Furthermore, the distribution of O^{8+}/p is consistent with the OVIII distribution across different solar wind types, displaying similar relative differences between Outlier and Streamer winds. In contrast, the distributions of other ion line fluxes, such as OVII, CVI, CV, and MgXI, differ significantly from the corresponding ion abundances from ACE (O^{7+}/p, C^{6+}/p, C^{5+}/p, Mg^{11+}/p). This discrepancy may be due to the fact that emissions at these energies can show greater spatial variability across the sky background compared to OVIII (Henley & Shelton, 2012), and due to the fact these lines are weaker than OVII. Therefore, future work should focus on refining the SWCX model used for spectral analysis of the XMM-Newton data.

Finally, we investigated whether the aid of XMM-Newton data can improve predictions of solar wind type, in the absence of heavy ion measurements from ACE/SWICS, as these measurements are no longer provided by any solar wind monitor. Our data set comprises a combination of solar wind proton properties from ACE/SWEPAM (proton density, velocity, flux, and temperature), IMF from ACE/MAG and XMM-derived features (lightcurve deviation and line fluxes from OVIII, OVII, CVI, CV and MgXI). We predicted solar wind types employing a Random Forest (Breiman, 2001) classifier. XMM-Newton data improves the performance of the Random Forest Classifier model to predict solar wind types when blind folded from ACE ion measurements. When using only ACE data, the macro-averaged F1 score, a measure that balances precision and recall, achieves a value of 0.55 with a standard deviation of 0.12. By incorporating XMM-Newton features into the model, the macro-averaged F1 score increases to 0.80 with a reduced standard deviation of 0.06. This notable improvement underscores the value of XMM-Newton data in enhancing the accuracy and reliability of solar wind type predictions, particularly when direct measurements of heavy ions are unavailable. Predicting solar wind types allows us to estimate ranges for solar wind composition. This approach is especially relevant for fully ionized ions, which lack excitation or emission, and can otherwise only be observed through direct in situ measurements.

Although XMM-Newton provides valuable insights into the solar wind type approaching Earth, this information is limited to when its line-of-sight traverses the magnetosheath. Additionally, XMM-Newton cannot provide detailed composition abundances of the heavy ions in the solar wind due to its moderate spectral resolution. Continuous monitoring of the solar wind ion composition is crucial for interpreting solar-terrestrial interactions, particularly the SWCX signal. This monitoring will be essential for analyzing upcoming data from the SMILE mission, a joint ESA-CAS collaboration, as well as NASA's LEXI mission. Both missions aim to capture large-scale magnetopause dynamics via soft X-ray imaging. Additionally, this analysis will be crucial for future X-ray missions currently in development or under study, for example, Athena, LEM, and GEO-X. Due to the declining performance of the solar wind ion composition spectrometer onboard ACE since 2011, it is crucial for the solar-terrestrial physics community to include high-cadence spectrometers within new missions. The recently proposed Elfen mission aims to address this data gap using a 16U CubeSat in a circular orbit 12 Earth radii around Earth (Carter et al., 2023). The CubeSat will carry two instruments: T-FIPS, a time-of-flight, energy ion mass spectrometer that measures heavy ions, developed from the FIPS instrument used on NASA's MESSENGER mission, and a magnetometer, MAGIC, to finally determine the plasma conditions inside and outside the Earth's magnetosphere encountered by the Elfen spacecraft.

Data Availability Statement

The ACE solar wind composition data were downloaded from CDAWeb at https://cdaweb.gsfc.nasa.gov/. The Random Forest Classifier was taken from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html. Sunspot number data was taken from NOAA at www.swpc.noaa.gov/products/solar-cycle-progression.

Acknowledgments
SN is supported by the University of Leicester's Future-100 Scholarship scheme. JAC and SN are supported by Royal Society Grant DHF\R1\211068. SFS is funded through the United Kingdom Space Agency, Grant RP16G1359. SEM is supported by the Natural Environment Research Council (NERC), UK, Grant NE/W006766/1. STL was supported in part by NASA contract 80NSSC23K0542. L.Z. acknowledges support from NASA Grants 80NSSC21K0579 and 80NSSC22K1015, NSF SHINE Grant 2229138, and NSF Early Career Grant 2237435. This research used the ALICE High Performance Computing facility at the University of Leicester. The authors are grateful to the National Space Science Data Centre for the ACE and OMNI data, and to the National Oceanic and Atmospheric Administration (NOAA) for the sunspot number data.

References

Bame, S., Asbridge, J., Feldman, W., & Kearney, P. (1974). The quiet corona: Temperature and temperature gradient. *Solar Physics*, *35*(1), 137–152. https://doi.org/10.1007/bf00156963

Bhardwaj, A. (2006). X-ray emission from Jupiter, Saturn, and Earth: A short review. In A. Bhardwaj (Ed.), *Advances in geosciences, volume 3: Planetary science (ps)* (Vol. 3, pp. 215–230). https://doi.org/10.1142/9789812707192_0021

Bodewits, D., Christian, D. J., Torney, M., Dryer, M., Lisse, C. M., Dennerl, K., et al. (2007). Spectral analysis of the Chandra comet survey. *A&A*, *469*(3), 1183–1195. https://doi.org/10.1051/0004-6361:20077410

Branduardi-Raymont, G. (2022). X-ray views of our solar system. *Astronomische Nachrichten*, *343*(4), e210101. https://doi.org/10.1002/asna.20210101

Branduardi-Raymont, G., Wang, C., Escoubet, C., Adamovic, M., Agnolon, D., Berthomier, M., et al. (2018). SMILE definition study report (Vol. *1*). https://doi.org/10.5270/esa.smile.definition_study_report-2018-12

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Carter, J. A., Samara-Ratna, P., Lepri, S. T., Raines, J. M., Vitkova, A., Brown, P., et al. (2023). Elfen: A CubeSat mission to measure heavy ions on both the dayside and nightside of Earth's magnetosphere. In *AGU Fall Meeting 2023. San Francisco (Califonia), United States*. Retrieved from https://insu.hal.science/insu-04351850

Carter, J. A., Sembay, S., & Read, A. M. (2011). Identifying XMM-Newton observations affected by solar wind charge exchange - Part II. *A&A*, *527*, A115. https://doi.org/10.1051/0004-6361/201015817

Cravens, T. E. (1997). Comet Hyakutake x-ray source: Charge transfer of solar wind heavy ions. *Geophysical Research Letters*, *24*(1), 105–108. https://doi.org/10.1029/96GL03780

Cravens, T. E., Robertson, I. P., & Snowden, S. L. (2001). Temporal variations of geocoronal and heliospheric X-ray emission associated with the solar wind interaction with neutrals. *Journal of Geophysical Research*, *106*(A11), 24883–24892. https://doi.org/10.1029/2000JA000461

Cucho-Padin, G., & Waldrop, L. (2019). Time-dependent response of the terrestrial exosphere to a geomagnetic storm. *Geophysical Research Letters*, *46*(21), 11661–11670. https://doi.org/10.1029/2019GL084327

Freyberg, M. J. (1998). On the zero-level of the soft X-ray background. In D. Breitschwerdt, M. J. Freyberg, & J. Truemper (Eds.), *Iau colloq* (Vol. 506, pp. 113–116). The Local Bubble and Beyond. https://doi.org/10.1007/BFb0104704

Gloeckler, G., & Geiss, J. (1989). The abundances of elements and isotopes in the solar wind. In C. J. Waddington (Ed.), *Cosmic abundances of matter* (Vol. 183, pp. 49–71). AIP. https://doi.org/10.1063/1.37985

Gloeckler, G., Geiss, J., Balsiger, H., Bedini, P., Cain, J. C., Fischer, J., et al. (1992). The solar wind ion composition spectrometer. *Astronomy and Astrophysics Supplement Series*, *92*(2), 267–289.

Henke, T., Woch, J., Schwenn, R., Mall, U., Gloeckler, G., von Steiger, R., et al. (2001). Ionization state and magnetic topology of coronal mass ejections. *Journal of Geophysical Research*, *106*(A6), 10597–10613. https://doi.org/10.1029/2000JA900176

Henley, D. B., & Shelton, R. L. (2012). An XMM-Newton survey of the soft X-ray background. II. An all-sky catalog of diffuse O VII and O VIII emission intensities. *ApJS*, *202*(2), 14. https://doi.org/10.1088/0067-0049/202/2/14

Hundhausen, A. J. (1972). Coronal expansion and solar wind. *Physics and Chemistry in Space*. https://doi.org/10.1007/978-3-642-65414-5

Hundhausen, A. J., Gilbert, H. E., & Bame, S. J. (1968). Ionization state of the interplanetary plasma. *Journal of Geophysical Research (1896-1977)*, *73*(17), 5485–5493. https://doi.org/10.1029/JA073i017p05485

Jansen, F., Lumb, D., Altieri, B., Clavel, J., Ehle, M., Erd, C., et al. (2001). XMM-Newton observatory. I. The spacecraft and operations. *Astronomy & Astrophysics*, *365*(1), L1–L6. https://doi.org/10.1051/0004-6361:20000036

Koutroumpa, D. (2024). Solar wind ion charge state distributions and compound cross sections for solar wind charge exchange x-ray emission. *Earth and Planetary Physics*, *8*(1), 1–14. https://doi.org/10.26464/epp2023056

Lallement, R. (2004). The heliospheric soft X-ray emission pattern during the ROSAT survey: Inferences on Local Bubble hot gas. *Astronomy & Astrophysics*, *418*(1), 143–150. https://doi.org/10.1051/0004-6361:20040059

McComas, D. J., Bame, S. J., Barker, P., Feldman, W. C., Phillips, J. L., Riley, P., & Griffee, J. W. (1998). Solar wind electron proton alpha monitor (SWEPAM) for the advanced composition explorer. *Space Science Reviews*, *86*(1/4), 563–612. https://doi.org/10.1023/A:1005040232597

Neugebauer, M. (1994). Observations of the solar wind from coronal holes. *Space Science Reviews*, *70*(1–2), 319–330. https://doi.org/10.1007/BF00777887

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Rakhmanova, L., Riazantseva, M., & Zastenker, G. (2016). Correlation level between solar wind and magnetosheath plasma and magnetic field parameters. *Advances in Space Research*, *58*(2), 157–165. https://doi.org/10.1016/j.asr.2015.09.036

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), A09104. https://doi.org/10.1029/2004JA010598

Richardson, I. G., & Cane, H. V. (2010). Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996 - 2009): Catalog and summary of properties. *Solar Physics*, *264*(1), 189–237. https://doi.org/10.1007/s11207-010-9568-6

Robertson, I. P., & Cravens, T. E. (2003). X-ray emission from the terrestrial magnetosheath. *Geophysical Research Letters*, *30*(8), 1439. https://doi.org/10.1029/2002GL016740

Samsonov, A., Sembay, S., Read, A., Carter, J. A., Branduardi-Raymont, G., Sibeck, D., & Escoubet, P. (2022). Finding magnetopause standoff distance using a soft x-ray imager: 2. Methods to analyze 2-d x-ray images. *Journal of Geophysical Research: Space Physics*, *127*(12), e2022JA030850. https://doi.org/10.1029/2022JA030850

Sembay, S., Alme, A. L., Agnolon, D., Arnold, T., Beardmore, A., Margeli, A. B. B., et al. (2024). The soft x-ray imager (SXI) on the smile mission. *Earth and Planetary Physics*, *8*(1), 5–14. https://doi.org/10.26464/epp2023067

Smith, C. W., L'Heureux, J., Ness, N. F., Acuña, M. H., Burlaga, L. F., & Scheifele, J. (1998). The ACE magnetic fields experiment. *Space Science Reviews*, *86*(1/4), 613–632. https://doi.org/10.1023/A:1005092216668

Stone, E. C., Frandsen, A. M., Mewaldt, R. A., Christian, E. R., Margolies, D., Ormes, J. F., & Snow, F. (1998). The advanced composition explorer. *Space Science Reviews*, *86*(1/4), 1–22. https://doi.org/10.1023/A:1005082526237

Strüder, L., Briel, U., Dennerl, K., Hartmann, R., Kendziorra, E., Meidinger, N., et al. (2001). The European photon imaging camera on XMM-Newton: The PN-CCD camera. *Astronomy & Astrophysics*, *365*(1), L18–L26. https://doi.org/10.1051/0004-6361:20000066

Turner, M. J. L., Abbey, A., Arnaud, M., Balasini, M., Barbera, M., Belsole, E., et al. (2001). The European photon imaging camera on XMM-Newton: The MOS cameras. *Astronomy & Astrophysics*, *365*(1), L27–L35. https://doi.org/10.1051/0004-6361:20000087

Von Steiger, R., & Zurbuchen, T. H. (2016). Solar metallicity derived from in situ solar wind composition. *The Astrophysical Journal*, *816*(1), 13. https://doi.org/10.3847/0004-637X/816/1/13

Whittaker, I. C., Sembay, S., Carter, J. A., Read, A. M., Milan, S. E., & Palmroth, M. (2016). Modeling the magnetospheric x-ray emission from solar wind charge exchange with verification from xmm-Newton observations. *Journal of Geophysical Research: Space Physics*, *121*(5), 4158–4179. https://doi.org/10.1002/2015JA022292

Zhao, L., Landi, E., Lepri, S. T., & Carpenter, D. (2022). Depletion of heavy ion abundances in slow solar wind and its association with quiet Sun regions. *Universe*, *8*(8), 393. https://doi.org/10.3390/universe8080393

Zhao, L., Landi, E., Lepri, S. T., Gilbert, J. A., Zurbuchen, T. H., Fisk, L. A., & Raines, J. M. (2017a). On the relation between the in situ properties and the coronal sources of the solar wind. *ApJ*, *846*(2), 135. https://doi.org/10.3847/1538-4357/aa850c

Zhao, L., Landi, E., Lepri, S. T., Kocher, M., Zurbuchen, T. H., Fisk, L. A., & Raines, J. M. (2017b). An anomalous composition in slow solar wind as a signature of magnetic reconnection in its source region. *The Astrophysical Journal Supplement Series*, *228*(1), 4. https://doi.org/10.3847/1538-4365/228/1/4

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. https://doi.org/10.1029/2009GL039181