Helios `corefit` data product user guide

David Stansby
david.stansby14@imperial.ac.uk

September 14, 2017

1 Introduction

This document describes the Helios `corefit` data product. This is a new data set which contains estimates of number density, velocity, and temperatures of the proton core population in the solar wind created from systematic fitting of all the original Helios 3D distribution functions.

2 Data availability and documentation

The data set is freely available at `ftp://apollo.ssl.berkeley.edu/pub/helios-data/E1_experiment/New_proton_corefit_data_2017/`. If you do use any of the data in a publication, I would be grateful if you could let me know by emailing david.stansby14@imperial.ac.uk.

Data are split by probe, year, and day of year. The data are available as either *cdf* or ascii *csv* files with data for each day contained in a single file. Table 1 summarises the variables present in the data product.

The original 3D ion distribution function data are available on the Helios archive FTP server at `ftp://apollo.ssl.berkeley.edu/pub/helios-data/E1_experiment/helios_raw/`. Detailed information about the plasma instrumentation can be found at `ftp://apollo.ssl.berkeley.edu/pub/helios-data/E1_experiment`. The source code used to read in and fit the distribution functions is available at `http://doi.org/10.5281/zenodo.891405`. This code can be used to reproduce the dataset from the original distribution function files.

Parameter	Parameter label	Units	Note
Time	Time		
Fitting status	status		a
Instrument	Instrument		b
Magnetic field instrument	B instrument		c
Proton number density	n_p	cm^{-3}	
Proton x velocity	vp_x	$km \cdot s^{-1}$	
Proton y velocity	vp_y	$km \cdot s^{-1}$	
Proton z velocity	vp_z	$km \cdot s^{-1}$	
Proton perpendicular thermal speed	vth_p_perp	$km \cdot s^{-1}$	
Proton parallel thermal speed	vth_p_par	$km \cdot s^{-1}$	
Proton perpendicular temperature	Tp_perp	Kelvin	
Proton parallel temperature	Tp_par	Kelvin	
B_x	Bx	nT	
B_y	By	nT	
B_z	Bz	nT	
Magnetic field standard deviation	sigma B	nT	d
Sun-spacecraft distance	r_sun	AU	
Carrington longitude	clong	Degrees	
Carrington latitude	clat	Degrees	
Carrington rotation number	carrot		
Spacecraft-Earth angle	earth_he_angle	Degrees	

Table 1: Description of variables present in the `core_fit` data product.

---

a. Fitting status flags take the following values: 1: successful fit; 2: successful fit, magnetic field varies too much for reliable number densities and temperatures; 3: successful fit, no magnetic field data available; 4: failed fit

b. Ion instrument flags take the following values: 1: I1a ion instrument data; 2: I3 ion instrument data

c. Magnetic field instrument flags take the following values: 1: E3 magnetometer data used; 2: E2 magnetometer data used

d. Magnetic field standard deviation is calculated as $\sqrt{\sigma_{Bx}^2 + \sigma_{By}^2 + \sigma_{Bz}^2}$, with individual standard deviations taken over the interval that the ion distribution function was measured

	B available	**B** varies too much	**B** not available	Failed fits
`Status` variable value	1	2	3	4
`B_instrument` variable value	1 or 2	-1	-1	-1
Number of points	1,869,275	34,759	192,677	119,484
% points	84.3%	1.6%	8.7%	5.4%
Number density	x			
Velocity	x	x	x	
Temperatures	x			
Magnetic field	x			

Table 2: Summary of which fitted parameters are available depending on magnetic field availability

3 Fitting method

The parameters in the `corefit` dataset were generated by fitting bi-Maxwellian functions to the original 3D ion distribution functions. The distributions contain contributions from both protons and alpha particles. Because they are well separated in energy, it is easy to fit a distribution to just the protons. Each distribution was fitted using the following process:

1. If magnetic field data was available, an average magnetic field was calculated from values taken whilst the distribution function was measured. The distribution function was then rotated into the field aligned frame.

2. A 3D bi-Maxwellian fit was done of the following form (fit parameters are underlined):

$$f_{fit}\left(v_{\parallel}, v_{\perp 1}, v_{\perp 2}\right) = \underline{A} \cdot \exp - \left\{ \left(\frac{v_{\parallel} - \underline{u}_{\parallel}}{\underline{w}_{\parallel}}\right)^2 + \left(\frac{v_{\perp 1} - \underline{u}_{\perp 1}}{\underline{w}_{\perp}}\right)^2 + \left(\frac{v_{\perp 2} - \underline{u}_{\perp 2}}{\underline{w}_{\perp}}\right)^2 \right\} \quad (1)$$

The 6 fit parameters are amplitude ($\underline{A}$), 3 velocities ($\underline{u}_{\parallel}$, $\underline{u}_{\perp 1}$, $\underline{u}_{\perp 2}$), and 2 thermal speeds ($\underline{w}_{\perp}$, $\underline{w}_{\parallel}$). The fit was done using least squares minimisation of the residuals calculated from ($f_{data} - f_{fit}$).

3. The number density was calculated from

$$n = \underline{A} \cdot \pi^{3/2} \underline{w}_{\perp} \underline{w}_{\perp} \underline{w}_{\parallel} \quad (2)$$

and two temperatures from

$$T_{\perp/\parallel} = \frac{m_p \underline{w}_{\perp/\parallel}^2}{2k_B} \quad (3)$$

If no magnetic field was available in step 1, steps 2 and 3 still took place, but the thermal speeds and subsequently the number density were ignored. If the magnetic field varies too much during the 32 seconds needed to measure a full distribution the number density and temperatures are unreliable and were thus also ignored.

Table 2 summarises which variables are available in each case. Figure 1 shows an example of the original distribution function along with the fit.
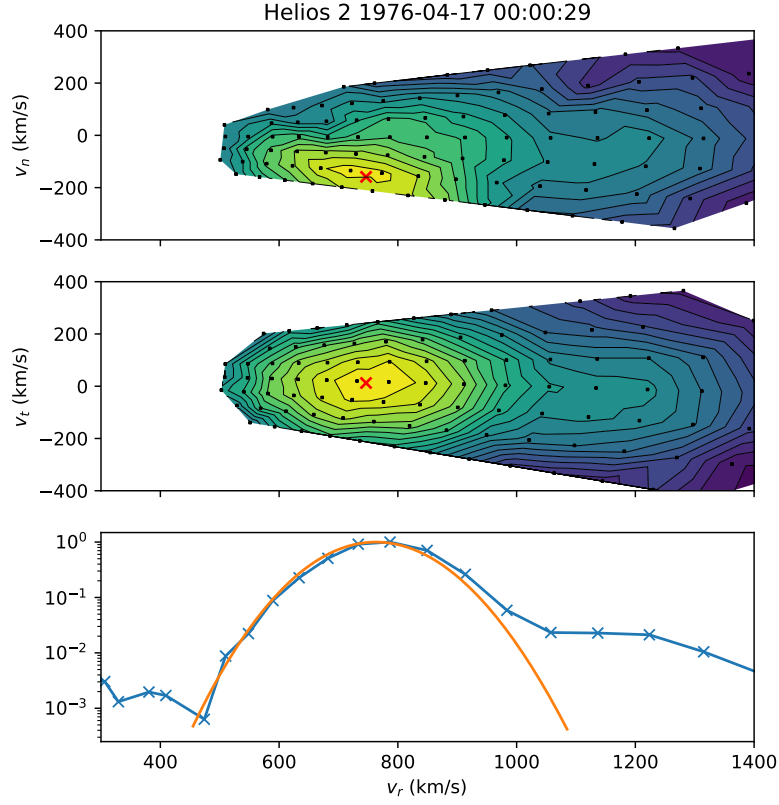
Figure 1: Example of distribution function data and corresponding fit. Top panel shows reduced 2D distribution function in the RN plane of an RTN co-ordinate system. Middle panel shows reduced distribution function in the RT plane. Black dots show positions in velocity space where the distribution function was sampled. Red crosses show the bulk velocity estimate from the fit, which agree well with the data. Bottom panel shows reduced 1D distribution function in blue, with the fit in orange. The fit agrees very well with the data, and is not sensitive to the small proton beam present between 900 to 1100 km/s or the alpha particle population present between 1100 to 1400 km/s.
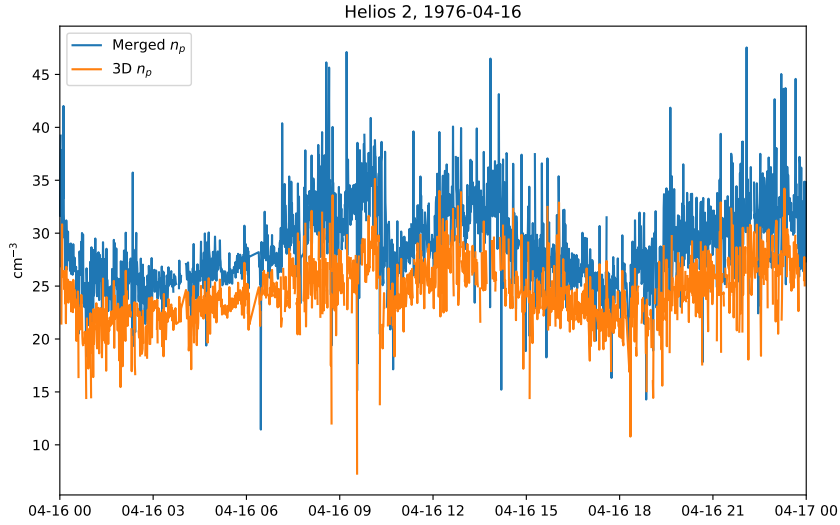
Figure 2: One day comparison of `merged` proton number density (blue) and `corefit` number density (orange). x-axis ticks are every 3 hours with the entire x-axis spanning 1 day, and formatted MM-DD HH.

4 Differences between `corefit` and `merged` data sets

The `merged` data set[1] is the only other freely available data set with estimates of proton bulk parameters. These parameters were calculated from numerical moments of reduced 1D distribution functions. This means that the `merged` data does not discriminate between the proton core and proton beam population. Differences between each parameter are as follows:

4.1 Number density

The number density in the `merged` data set contains contributions from the proton beam, so is systematically higher than the `corefit` number density. The difference is typically around 10%, but can be as high as 40% at times. A time series comparison is shown in figure 2.

4.2 Velocity

The radial component of velocity is systematically higher in the `merged` data set compared to the `corefit` values, due to the presence of the proton beam. The azimuthal and polar components are not affected by this and the two data sets contain similar values. A time series comparison is shown in figure 3.

4.3 Temperature

The `merged` data set contains only one proton temperature value. This value was calculated from the 1D distribution function which means it contains variable contributions from the
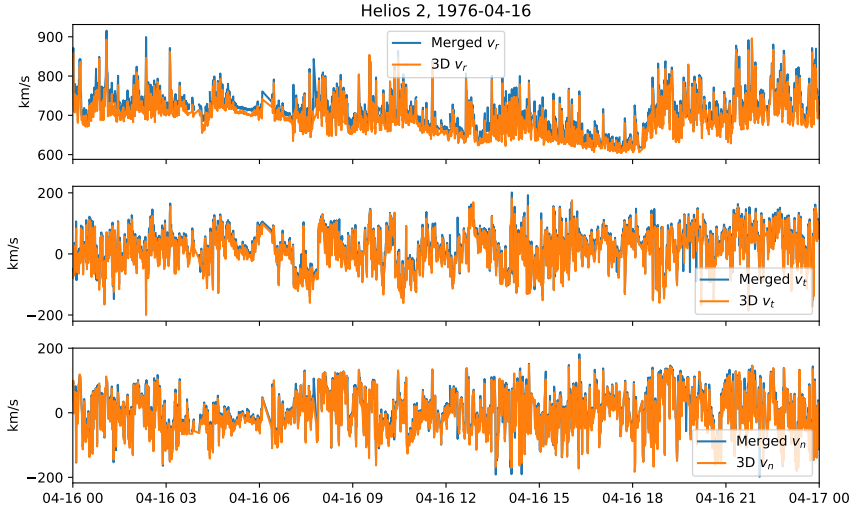
---

[1]Available at `ftp://cdaweb.gsfc.nasa.gov/pub/data/helios/helios1/merged/` and `ftp://cdaweb.gsfc.nasa.gov/pub/data/helios/helios2/merged/`

Figure 3: One day comparison of `merged` proton velocity (blue) and `corefit` proton velocity (orange) in an RTN co-ordinate system. x-axis ticks are every 3 hours with the entire x-axis spanning 1 day, and formatted MM-DD HH.

true parallel and perpendicular temperatures of the protons. The perpendicular and parallel temperatures in the `corefit` data set are much more accurate estimates of the temperatures of the proton core. A time series comparison is shown in figure 4.

The total temperature, which can be calculated from the parallel and perpendicular temperatures via.

$$T = \frac{2T_\perp + T_\parallel}{3} \tag{4}$$

is therefore much more accurate in the `corefit` data set. The difference between the `merged` temperature and the `corefit` temperature can be as big as a factor of 5 (see figure 4, middle panel).

5   A note on timestamps

The timestamps in this data set were taken directly from the timestamps given in the filenames of the individual distribution function files. The distributions were originally recorded with a nominal cadence of 40.5 seconds, but it is believed anything after the decimal point in the timestamp has been dropped. This means that the time difference between consecutive time stamps goes like $0, 40, 81, 121, 162...$, whereas the true time differences were $0, 40.5, 81, 121.5, 162....$
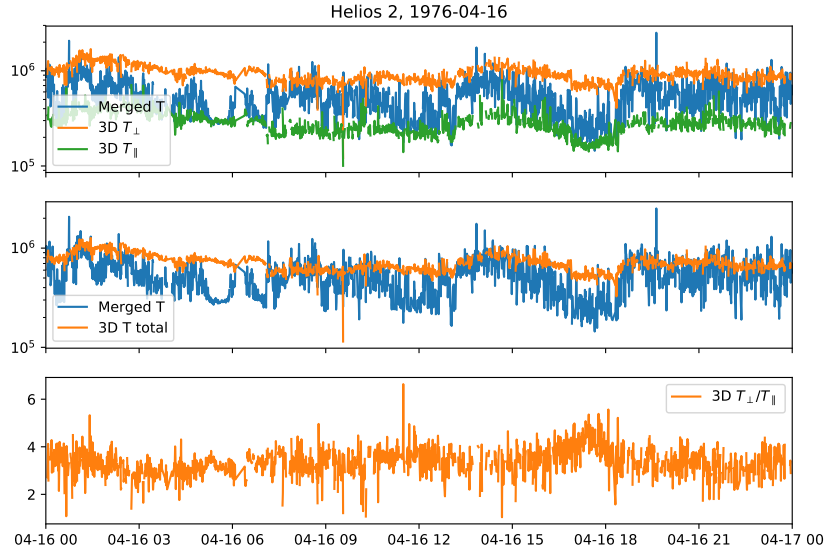
Figure 4: One day comparison of `merged` proton temperature and `corefit` proton temperatures. Top panel shows `merged` proton temperature (blue), `corefit` perpendicular temperature (orange), and `corefit` parallel temperature (green). Middle panel shows `merged` proton temperature (blue) and `corefit` proton temperature (orange). Bottom panel shows `corefit` temperature anisotropy (orange). x-axis ticks are every 3 hours with the entire x-axis spanning 1 day, and formatted MM-DD HH.

A Acknowledgements