Chapter 4
Preparing Spacecraft or Ground-Based Data for Fitting

This chapter describes the spacecraft or ground-based data and the methodology for preparing and selecting the real data for the fitting program.

4.1 Data Selection

There are spacecraft orbiting the Earth and some particle detectors set on the ground that were built for studying space weather, solar energetic particles, including the mechanisms of the solar injection, the interplanetary magnetic field, or the solar wind. Some spacecraft or ground-based detectors detect the energetic particles or space plasma and then the data are publicly shown on a website. We can choose the spacecraft or instrument that has interesting data for our analysis.

In this work we are interested in spacecraft or ground-based data that give details about the distribution profile of solar energetic particles in time. Sometimes, there are gaps in the data, or the data can not immediately be used in our work. For example, we want to know details about the time, flux and its uncertainty, but spacecraft data typically give the data of time, flux and counts, so we must manipulate the data files. Some points that we should consider in the selected data are:

1. What is the particle? If we know the type of the particle, we will know its mass. The mass of the particle is an input value for the simulation program.

2. What is the kinetic energy of the selected particles? We fit the spacecraft or ground-based data with the results from simulations for the same energy. The fitting results report the best mean free path for each energy range.

3. The data are not suitable for analysis if there is a strong, time-dependent influence from the interplanetary medium such as the effect from a shock wave. A shock wave can affect the data by freshly accelerating particles (which is not currently considered in our simulations) or causing a sudden variations in the data.

4. The profile of the chosen data smoothly increases and decays in time compared with the background level data. Such a time dependence should be explained by transport effects, as treated in our code, and is thus suitable for analysis.

5. The interplanetary magnetic field during the time period of interest is stable, or there is little variation in the range of the magnetic field data.

6. The background value of the data of interest is the average value of the range of the data before the onset of the SEP event.

7. We cut some data in which the intensity or anisotropy×intensity is equal to zero, where the intensity is the average flux of particle detected from all directions. "Isotropy" implies a state in which the flux is equal in every direction; an "anisotropic" distribution is not isotropic. More specifically, for solar energetic particles, the (dipole) anisotropy times the intensity is the first-order term in a Legendre expansion in the direction outward from the Sun along the magnetic field, and the anisotropy is this quantity divided by the intensity.

The data are compared with the results of the simulation program. The selected spacecraft or ground-based data and the results from the simulation program must be for the same element, energy interval, and time period. The spacecraft or ground-based data we use are in the form of three columns: time, intensity or intensity×anisotropy of particles, and the uncertainty. We use the spacecraft or ground-based data that correspond to our conditions to be the data for fitting in comparison with our simulation results.

4.2 Additional Information about Events of Interest

In addition to the data file for fitting, which we prepared from the spacecraft or ground-based data, it is useful to find additional information such as:

1. The solar wind velocity measured at a spacecraft is used in the simulation program. We can find the solar wind speed from the Solar Wind Ion Composition Spectrometer (SWICS) and the Solar Wind Ion Mass Spectrometer (SWIMS) on the Advanced Composition Explorer (ACE) spacecraft. If the solar wind speed was not measured at the time period of interest, we have to input a typical value, such as 400 km/s, or a value roughly estimated from neighboring time periods.

2. We can check for shock effects on the event of interest by examining the magnetic field data. If the interplanetary magnetic field lines suddenly change, we have to watch out for the influence of a shock. This is an important condition for deciding whether to analyze a given data set.

3. If we find more information about the solar event, such as the X-ray intensity, the duration of the event, the start time, the CME data at, before, and after the solar event of interest occurred, this will be useful in preparing the simulations and understanding more details of the fitting results.

All of these points are very helpful for the initial value preparation for simulations which correspond to the selected solar event. Every time we select a solar event for fitting, we should consider this information, most of which can be found by means of the World Wide Web.

4.3 Uncertainty of Spacecraft Data: Taking Interplanetary Fluctuations into Account

The spacecraft data are usually provided along with uncertainties, which are statistical errors. The statistical uncertainty was derived from the flux data, with the fractional uncertainty as $1/\sqrt{N}$, where N is the number of particles counted in the averaging period. We found that sometimes the statistical uncertainty does not correspond to the actual fluctuations. As an example, we show the spacecraft data for oxygen at 15.6-21.0 MeV/nucleon on Nov. 6, 1997, with statistical uncertainties as shown in Figure 4.1. These data come from http://www.srl.caltech.edu/ACE/ASC/level2/sis_l2desc.html. We found that near the peak, the data have statistical uncertainties that are much lower than that the level of fluctuations. We believe that this is because the fluctuating connection of the turbulent interplanetary magnetic field between the Sun and the Earth affects the data. [We can check the interplanetary magnetic field profile of the selected data from the Magnetometer (MAG) instrument on the Advanced Composition Explorer (ACE) spacecraft (http://www.srl.caltech/ACE).] Taking account of the uncertainty of the data due to interplanetary fluctuations should give more accurate fitting results. This idea was developed in discussions with Professor Dr. Glenn Mason, Dr. Joseph Dwyer, Dr. Mihir Desai, and my advisor.

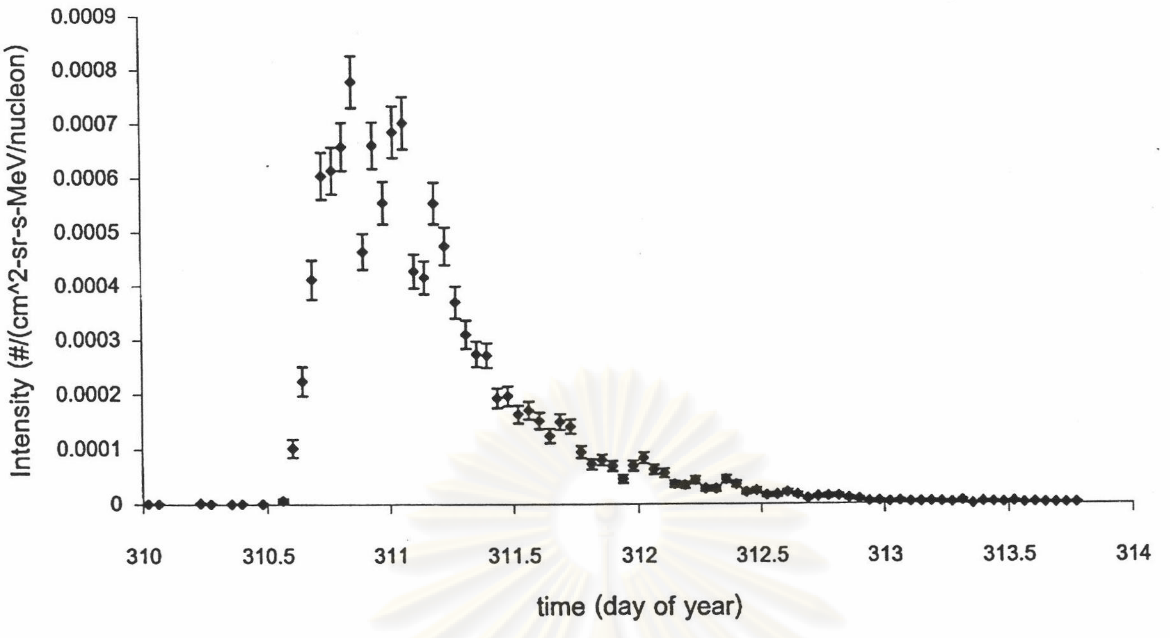

Figure 4.1: An example of spacecraft data for oxygen with the statistical uncertainties.

We have developed a new technique for preparing the uncertainty of the spacecraft data, taking non-statistical fluctuations into account. The idea of interplanetary fluctuations is that some magnetic flux tubes contain many particles, but some may have only a few. We cannot control the uncertainty in the interplanetary fluctuations, so we calculated the non-statistical uncertainties from the standard deviation of selected data from a trendline (an example for calculating the total uncertainties is in Appendix A). The new total uncertainty of the data combines statistical and interplanetary fluctuation errors. We use the notation:

σ_{tot} the total uncertainty in intensity due to both statistical uncertainty and interplanetary fluctuations,

σ_{stat} the statistical error of each intensity (flux) or counting statistics,

σ_{ipf} the standard derivation of intensity due to the interplanetary fluctuations, The procedure of finding the total uncertainty is:

1. We identify the maximum point of the data and a reference time period. The number of the reference points depends on the user and the event or data set. In this work we choose six points including and after the maximum point of the data as the reference points, because the trend line of these 6 points can be taken to be a straight line.

2. We plot the intensity of particles versus time for the reference time period to find the formula for the trend line of the reference data points.

3. We find the statistical error in the reference interval ($\sigma_{stat,0}$) from

$$\sigma^2_{stat,0} = \frac{1}{N}\sum_i \sigma^2_{stat,i}, \tag{4.1}$$

where N is the number of data points.

4. We find the standard deviation from the trendline in the reference interval ($\sigma_{tot,0}$) from

$$\sigma^2_{tot,0} = \frac{1}{N-2}\sum_i [I_i - \hat{I}(t_i)]^2, \tag{4.2}$$

where N is the number of data points, I_i is the intensity of particles at each time, $\hat{I}(t_i)$ is the intensity of particles corresponding to the trendline equation.

5. If $\sigma_{tot,0} < \sigma_{stat,0}$, we set the total uncertainty, σ_{tot}, equal to the statistical uncertainty, σ_{stat}.

6. Otherwise, we use the total uncertainty (σ_{tot}) as

$$\sigma^2_{tot} = \sigma^2_{stat} + \sigma^2_{ipf}, \tag{4.3}$$

assuming that statistical and interplanetary fluctuations are independent. For the reference data points, this gives

$$\begin{aligned} \sigma^2_{ipf,0} &= \sigma^2_{tot,0} - \sigma^2_{stat,0} && (4.4)\\ \sigma^2_{ipf,0} &= \frac{1}{N-2}\sum_i [I_i - \hat{I}(t_i)]^2 - \frac{1}{N}\sum_i \sigma^2_{stat,i}. && (4.5) \end{aligned}$$

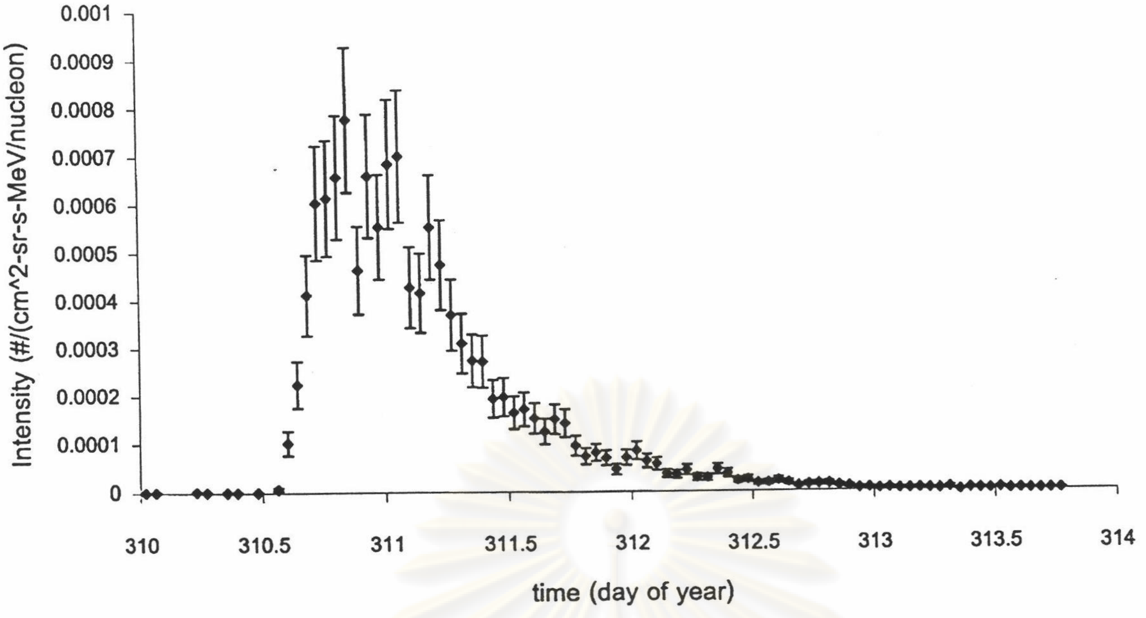

Figure 4.2: The spacecraft data for oxygen with the new uncertainties.

and for other data points we use

$$\sigma_{ipf} = \sigma_{ipf,0}\frac{I}{I_0}, \tag{4.6}$$

where I is the particle intensity at each point, and I_0 is the mean particle intensity in the reference interval.

From Figure 4.2, the new uncertainties include the effect of interplanetary fluctuations, not only statistical uncertainties as in Figure 4.1, giving what is clearly a better estimation of the level of fluctuations of the data. These new uncertainties should lead to accurate fitting.

4.4 Sources of Data

In this work, the spacecraft or ground-based data of interest come from 3 instruments.

1. A worldwide neutron monitor network is operated by institutions in

various nations. These neutrons are secondary particles generated by the collisions of ions (mainly protons) at the top of the atmosphere. Thus each neutron monitor measures the number of energetic particles coming from an asymptotic direction, which is determined by the geomagnetic field. We use data from various neutron monitor stations, including McMurdo DM, South Pole, Mawson, Sanae, and Terre Adelie in Antarctica; Thule in Greenland; Inuvik in Canada; and Tixie Bay and Apatity in Russia. The data of interest were sent to us by John W. Bieber and his colleagues at the Bartol Research Institute, University of Delaware. These data combine the intensity and anisotropy as a function of time.

2. The Wind spacecraft, which was launched on November 1, 1994 by a Delta rocket from Cape Canaveral, Florida. It was built for studying the space environment around the Earth, and has several instruments such as the Magnetic Field Instrument (MFI), the Energetic Particle Acceleration, Composition and Transport (EPACT) experiment, and the Solar Wind Experiment (SWE). The data of interest were sent to us by Wolfgang Dröge, Bartol Research Institute, University of Delaware, Newark, USA. These data are the intensity and anisotropy as a function of time.

3. The Solar Isotope Spectrometer (SIS), which is one instrument in the Advanced Composition Explorer (ACE) spacecraft. SIS was built by the California Institute of Technology, NASA Goddard Space Flight Center, and NASA Jet Propulsion Laboratory. The objectives of SIS are measurements of the isotropic composition of nuclei from He to Ni over the energy range from ~10 to ~100 MeV/nucleon (http://www.srl.caltech.edu/ACE/CRIS_SIS/sis.html). We can download the data of interest by choosing an element, energy band, and

time period. These data combine the intensity (flux) as a function of time and their counts (N), from which we can calculate the statistical uncertainties using $(\text{flux})/\sqrt{N}$.

In this work, we also consider other instruments on ACE for studying other effects on the event of interest, such as the details of the solar wind from the Solar Wind Electron, Proton and Alpha Monitor, SWEPAM, or the magnetic profile during a solar event from the MAG instrument.